#### (19) World Intellectual Property Organization International Bureau



# ! # 1/4 | 1/1/| # 1/1 | # 1/1/| | 1/1/| | 1/1/| | 1/1/| | 1/1/| | 1/1/| | 1/1/| | 1/1/| | 1/1/| | 1/1/| | 1/1/

(43) International Publication Date 26 February 2004 (26.02.2004)

### (10) International Publication Number WO 2004/016637 A1

(51) International Patent Classification7:

C07H 21/00

(21) International Application Number:

PCT/KR2003/001655

(22) International Filing Date: 14 August 2003 (14.08.2003)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

60/402,905

14 August 2002 (14.08.2002) US

60/403,651

16 August 2002 (16.08.2002)

- (71) Applicant (for all designated States except US): LG LIFE SCIENCES LTD. [KR/KR]; LG Twin Tower, East Tower, 20, Yoido-Dong, Youngdungpo-Ku, Seoul 150-010 (KR).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): KOH, Sang Seok [KR/KR]; R & D Park, LG Life Sciences Ltd., 104-1, Moongi-dong, Yuseong-gu, Taejeon 305-380 (KR). LIU, Qing [US/US]; 708 Quince Orchard Road, Gaithersburg, Maryland, MD 20878 (US). CHUNG, Hvun-Ho [KR/KR]; R & D Park, LG Life Sciences Ltd., 104-1, Moongi-dong, Yuseong-gu, Taejeon 305-380 (KR). ZENG, Wen [US/US]; 708 Quince Orchard Road, Gaithersburg, Maryland, MD 20878 (US). LEE, Bogman [KR/KR]; R & D Park, LG Life Sciences Ltd., 104-1, Moongi-dong, Yuseong-gu, Taejeon 305-380 (KR).

SONG, Si Young [KR/KR]; College of Medicine, Yonsei University, 134, Sinchon-dong, Seodaemun-gu, Seoul 120-749 (KR).

- (74) Agent: CHOI, Kyu Pal; Halla Classic Building 4F., 824-11, Yeoksam-dong, Kangnam-ku, Seoul 135-080 (KR).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

#### Published:

with international search report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



(54) Title: GENE FAMILIES ASSOCIATED WITH LIVER CANCER

(57) Abstract: The invention relates generally to the changes in gene expression in hepatocellular carcinoma. The invention relates specifically to human genes that correspond to mRNA species that are differentially expressed in cancerous liver tissueand in cancerous neoplasms compared to non-cancerous liver tissue.

### GENE FAMILIES ASSOCIATED WITH LIVER CANCER

#### TECHNICAL FIELD

The invention relates generally to the changes in gene expression in liver tissue from cancer patients who concurrently suffer from cirrhosis or hepatitis. The invention specifically relates to human gene families which are differentially expressed in hepatic carcinoma tissue, compared to inflamed or cirrhotic liver tissue, and in other malignant neoplasms.

10

15

20

25

5

#### **BACKGROUND ART**

#### Liver Disease

Generally, liver disease is classified as a disorder that causes the liver to malfunction or cease functioning all together. Cirrhosis, for example, is a group of chronic liver diseases in which liver cells are damaged and then replaced with scar tissue, thereby decreasing the amount of normal liver tissue. While it is most often caused by alcohol abuse, patients with hepatitis infections and other biliary diseases can also develop cirrhosis. Chronic hepatitis-B infection, hepatitis-C infection, and cirrhosis have all been shown to have strong associations with primary liver cancer, although the mechanisms involved are still not fully understood (Wu et al., (2001) Oncogene 20:3674-3682). About 10-20% of chronic hepatitis-B infections result in primary liver cancer. Other factors such as alcohol consumption, poor nutrition and aflatoxins are also linked to the development of primary liver cancer and cirrhosis.

Cirrhosis of the liver is characterized by widespread nodules combined with fibrosis. Damaged or dead liver cells are replaced by fibrous scar tissue, which to leads to fibrosis. Liver cells regenerate in an abnormal pattern, producing nodules surrounded by fibrous tissue. The fibrosis and nodule formation cause distortion and blockage of the liver's structural components, causing impaired blood flow and biochemical function.

10

15

20

25

30

T/KR2003/001655

In patients displaying overt symptoms, diagnosis of cirrhosis is usually easy, but cirrhosis may be difficult to detect in its early stages. Subtle changes occurring in the early stages include red palms, red spots on the upper body that blanch, hypertrophy of the parotid glands, fibrosis of the tendons in the palms and gynecomastia. X-rays and radioactive tracer tests may be effective, but diagnosis must often be by liver biopsy.

In contrast to the underlying pathology of cirrhosis, in primary liver cancer, liver cells become abnormal, grow out of control and form malignant tumors. This disease is also called hepatocellular carcinoma (HCC) or malignant hepatoma. Cancer that spreads to the liver from another part of the body as a result of metastasis is not the same disease. HCC is difficult to detect at an early stage because the symptoms are not specific. They include loss of appetite and weight, fever, fatigue and weakness. As the cancer progresses, pain may develop in the upper abdomen, extending to the back and right shoulder. Swelling or a palpable mass may also be present in the upper abdomen, along with jaundice and darkened urine. When the cancer metastasizes, it typically targets the lungs and brain.

Diagnosis of HCC may be made by blood tests, in particular, tests for tumor markers such as alpha-fetoprotein. About 50-70% of HCC patients show elevated levels of alpha-fetoprotein. Additional diagnostic methods include non-radioactive imaging (abdominal or chest x-rays, angiograms, CT scans and MRIs), liver scans using radioactive materials and liver biopsies. Treatment of HCC is often not successful, because detection is often too late, but methods include surgical removal of the cancer, chemotherapy and radiation, alone or in combination. Although HCC is not very common in the United States, it is very prevalent in parts of Asia and Africa, largely due to the higher incidence of infection with hepatitis viruses (http://cis.nci.nih.gov/; http://cancer.med.upenn.edu/disease/liver/intro liver.html).

#### Molecular Changes in Liver Disease

Little is known about the molecular changes in liver cells associated with the development and progression of liver disease. Accordingly, there exists a need for the investigation of the changes in gene expression levels as well as the need for the

identification of new molecular markers associated with the development and progression of liver disease. Furthermore, if intervention is expected to be successful in halting or slowing down liver disease, means of accurately assessing the early manifestations of cirrhosis or HCC need to be established. Likewise, the development of therapeutics to prevent or stop the progression of liver disease relies on the identification of genes responsible for the cancerous transformation of liver cells and the growth of cancerous liver cells or the induction of tissue damage and scar formation associated with cirrhosis.

#### DISCLOSURE OF THE INVENTION

10

15

20

25

5

The present invention is based on the discovery of new gene families, each designated LBFL302 and LBFL303, that are differentially expressed in hepatocellular carcinoma (HCC), compared to liver cirrhosis (LC) or chronic hepatitis (CH), and in other malignant neoplasms. The invention includes an isolated nucleic acid molecule selected from the group consisting of an isolated nucleic acid molecule comprising SEQ ID NO: 1, 3, 5 or 7, an isolated nucleic acid molecule encoding SEQ ID NO: 6 or 8, an isolated nucleic acid molecule that encodes a protein that is expressed in liver cancer and that exhibits at least about 95% nucleotide sequence identity over the entire contiguous sequence of SEQ ID NO: 5, an isolated nucleic acid molecule that encodes a protein that is expressed in liver cancer and that exhibits at least about 75% nucleotide sequence identity over the entire contiguous sequence of SEQ ID NO: 7, and an isolated nucleic acid molecule comprising the complement of any of the aforementioned nucleic acid molecules.

The present invention further includes the nucleic acid molecules operably linked to one or more expression control elements, including vectors comprising the isolated nucleic acid molecules. The invention further includes host cells transformed to contain the nucleic acid molecules of the invention and methods for producing a protein comprising the step of culturing a host cell transformed with a nucleic acid molecule of the invention under conditions in which the protein is expressed.

The invention further provides an isolated polypeptide selected from the group

10

15

20

consisting of an isolated polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 4, 6 or 8, an isolated polypeptide comprising a fragment of at least 10 amino acids of SEQ ID NO: 2 or 4, an isolated polypeptide comprising conservative amino acid substitutions of SEQ ID NO: 2 or 4 and an isolated polypeptide comprising naturally occurring amino acid sequence variants of SEQ ID NO: 2 or 4. Polypeptides of the invention also include polypeptides with an amino acid sequence having at least about 50%, 60%, 70% or 75% amino acid sequence identity with the sequence set forth in SEQ ID NO: 2 or 4, preferably at least about 80%, more preferably at least about 90-95%, and most preferably at least about 95-98% sequence identity with the sequence set forth in SEQ ID NO: 2 or 4, and a protein having at least about 95% amino acid sequence identity with SEQ ID NO: 6 or 8.

The invention further provides an isolated antibody or antigen-binding antibody fragment that specifically binds to a polypeptide of the invention, including monoclonal and polyclonal antibodies.

The invention further provides methods of identifying an agent which modulates the expression of a nucleic acid molecule encoding a protein of the invention, comprising: exposing cells which express the nucleic acid molecule to the agent; and determining whether the agent modulates expression of said nucleic acid molecule, thereby identifying an agent which modulates the expression of a nucleic acid molecule encoding the protein.

The invention further provides methods of identifying an agent which modulates the level of or at least one activity of a protein of the invention, comprising: exposing cells which express the protein to the agent; and determining whether the agent modulates the level of or at least one activity of said protein, thereby identifying an agent which modulates the level of or at least one activity of the protein.

The invention further provides methods of identifying binding partners for a protein of the invention, comprising the steps of exposing said protein to a potential binding partner; and determining if the potential binding partner binds to said protein, thereby identifying binding partners for the protein.

The present invention further provides methods of modulating the expression of

25

10

15

20

25

a nucleic acid molecule encoding a protein of the invention, comprising the step of administering an effective amount of an agent which modulates the expression of a nucleic acid molecule encoding the protein. The invention also provides methods of modulating at least one activity of a protein of the invention, comprising the step of administering an effective amount of an agent which modulates at least one activity of the protein of the invention.

The present invention further includes non-human transgenic animals modified to contain the nucleic acid molecules of the invention, or non-human transgenic animals modified to contain the mutated nucleic acid molecules such that expression of the encoded polypeptides of the invention is prevented.

The present invention also includes non-human transgenic animals in which all or a portion of a gene comprising all or a portion of SEQ ID NO: 1, 3, 5, 7 or 9 has been knocked out or deleted from the genome of the animal.

The invention further provides methods of diagnosing liver cancer and other cancers, comprising the steps of acquiring a tissue, blood, urine or other sample from a subject and determining the level of expression of a nucleic acid molecule of the invention or polypeptide of the invention.

The invention further includes compositions comprising a diluent and a polypeptide or protein selected from the group consisting of an isolated polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 4, 6, 8 or 10, an isolated polypeptide comprising a fragment of at least 10 amino acids of SEQ ID NO: 2, 4, 6, 8 or 10, an isolated polypeptide comprising conservative amino acid substitutions of SEQ ID NO: 2 or 4, naturally occurring amino acid sequence variants of SEQ ID NO: 2 or 4, an isolated polypeptide with an amino acid sequence having at least about 50%, 60%, 70% or 75% amino acid sequence identity with the sequence set forth in SEQ ID NO: 2 or 4, preferably at least about 80%, more preferably at least about 90-95%, and most preferably at least about 95-98% sequence identity with the sequence set forth in SEQ ID NO: 2 or 4, and a polypeptide having at least about 95% amino acid sequence identity with SEQ ID NO: 6, 8 or 10.

#### **BRIEF DESCRIPTION OF THE DRAWINGS**

Figure 1 is a hydrophobicity plot of the protein encoded by the open reading frame of LBFL302, variant BC4 (SEQ ID NO: 2). Analysis was performed according to the methods of Kyte-Doolittle and Goldman et al.

Figure 2 is a hydrophobicity plot of the protein encoded by the open reading frame of LBFL302, variant BC7 (SEQ ID NO: 4). Analysis was performed according to the methods of Kyte-Doolittle and Goldman *et al*.

10

20

30

5

- Figure 3 is a hydrophobicity plot of the protein encoded by the open reading frame of LBFL303, clone GE6 (SEQ ID NO: 6). Analysis was performed according to the methods of Kyte-Doolittle and Goldman *et al*.
- Figure 4 is a hydrophobicity plot of the protein encoded by the open reading frame of LBFL303, clone MB5 (SEQ ID NO: 8). Analysis was performed according to the methods of Kyte-Doolittle and Goldman et al.
  - <u>Figure 5</u> Figure 5 is a hydrophobicity plot of the protein encoded by the open reading frame of LBFL303, clone IE4 (SEQ ID NO: 10). Analysis was performed according to the methods of Kyte-Doolittle and Goldman *et al*.

#### BEST MODE FOR CARRYING OUT THE INVENTION

#### 25 I. General Description

The present invention is based in part on the identification of new gene families (LBFL302 and LBFL303) that are differentially expressed in cancerous human liver tissue, compared to inflamed or cirrhotic human liver tissue, and in other malignant neoplasms. These gene families correspond to the human cDNA of SEQ ID NOS: 1 and 3 (LBFL302), and SEQ ID NOS: 5, 7 and 9 (LBFL303).

10

15

20

25

30

The genes and proteins of the invention may be used as diagnostic agents or markers to detect liver cancer or to differentiate hepatic carcinoma from cirrhotic liver tissue in a sample. They can also serve as a target for agents that modulate gene expression or activity. For example, agents may be identified that modulate biological processes associated with tumor growth, including the hyperplastic process of liver cancer.

#### II. Specific Embodiments

#### A. The Proteins Associated with Liver Cancer

The present invention provides isolated proteins, allelic variants of the proteins, and conservative amino acid substitutions of the proteins. As used herein, the "protein" or "polypeptide" refers, in part, to a protein that has the human amino acid sequence depicted in SEQ ID NO: 2, 4, 6, 8 or 10. The terms also refer to naturally occurring allelic variants and proteins that have a slightly different amino acid sequence than that specifically recited above. Allelic variants, though possessing a slightly different amino acid sequence than those recited above, will still have the same or similar biological functions associated with these proteins.

As used herein, the family of proteins related to the human amino acid sequence of SEQ ID NO: 2, 4, 6, 8 or 10 refers to proteins that have been isolated from organisms in addition to humans. The methods used to identify and isolate other members of the family of proteins related to these proteins are described below.

The proteins of the present invention are preferably in isolated form. As used herein, a protein is said to be isolated when physical, mechanical or chemical methods are employed to remove the protein from cellular constituents that are normally associated with the protein. A skilled artisan can readily employ standard purification methods to obtain an isolated protein.

The proteins of the present invention further include insertion, deletion or conservative amino acid substitution variants of SEQ ID NO: 2, 4, 6, 8 or 10. As used herein, a conservative variant refers to alterations in the amino acid sequence that do not adversely affect the biological functions of the protein. A substitution, insertion or

• 5

10

15

20

25

30

deletion is said to adversely affect the protein when the altered sequence prevents or disrupts a biological function associated with the protein. For example, the overall charge, structure or hydrophobic/hydrophilic properties of the protein, in certain instances, may be altered without adversely affecting a biological activity. Accordingly, the amino acid sequence can be altered, for example to render the peptide more hydrophobic or hydrophilic, without adversely affecting the biological activities of the protein.

Ordinarily, the allelic variants, the conservative substitution variants, and the members of the protein family encoded by LBFL302 gene, will have an amino acid sequence having at least about 50%, 60%, 70% or 75% amino acid sequence identity with the sequence set forth in SEQ ID NO: 2 or 4, more preferably at least about 80%, even more preferably at least about 90-95%, and most preferably at least about 99 or 99.5% sequence identity. Further, those of the protein family encoded by LBFL303 gene, will have an amino acid sequence having at least about 50%, 60%, 70% or 75% amino acid sequence identity with the sequence set forth in SEO ID NO: 6, 8 or 10, more preferably at least about 80-90%, even more preferably at least about 91-94%, and most preferably at least about 95% or 98% sequence identity. Identity or homology with respect to such sequences is defined herein as the percentage of amino acid residues in the candidate sequence that are identical with SEQ ID NO: 2, 4, 6, 8 or 10, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent homology, and not considering any conservative substitutions as part of the sequence identity (see section B for the relevant parameters). Fusion proteins, or N-terminal, C-terminal or internal extensions, deletions, or insertions into the peptide sequence shall not be construed as affecting homology.

Thus, the proteins of the present invention include molecules having the amino acid sequence disclosed in SEQ ID NO: 2, 4, 6, 8 or 10; fragments thereof having a consecutive sequence of at least about 3, 4, 5, 6, 10, 15, 20, 25, 30, 35 or more amino acid residues of these proteins; amino acid sequence variants wherein one or more amino acid residues has been inserted N- or C-terminal to, or within, the disclosed coding sequence; and amino acid sequence variants of the disclosed sequence, or their fragments

10

15

20

25

30

as defined above, that have been substituted by at least one residue. Such fragments, also referred to as peptides or polypeptides, may contain antigenic regions, functional regions of the protein identified as regions of the amino acid sequence which correspond to known protein domains, as well as regions of pronounced hydrophilicity. The regions are all easily identifiable by using commonly available protein sequence analysis software such as MacVector (Oxford Molecular).

Contemplated variants further include those containing predetermined mutations by, e.g., homologous recombination, site-directed or PCR mutagenesis, and the corresponding proteins of other animal species, including but not limited to rabbit, mouse, rat, porcine, bovine, ovine, equine and non-human primate species, and the alleles or other naturally occurring variants of the family of proteins; and derivatives wherein the protein has been covalently modified by substitution, chemical, enzymatic, or other appropriate means with a moiety other than a naturally occurring amino acid (for example a detectable moiety such as an enzyme or radioisotope).

The present invention further provides compositions comprising a protein or polypeptide of the invention and a diluent. Suitable diluents can be aqueous or non-aqueous solvents or a combination thereof, and can comprise additional components, for example water-soluble salts or glycerol, that contribute to the stability, solubility, activity, and/or storage of the protein or polypeptide.

As described below, members of the family of proteins can be used: (1) to identify agents which modulate the level of or at least one activity of the protein, (2) to identify binding partners for the protein, (3) as an antigen to raise polyclonal or monoclonal antibodies, (4) as a therapeutic agent or target and (5) as a diagnostic agent or marker of liver cancer and other hyperplastic diseases.

#### B. Nucleic Acid Molecules

The present invention further provides nucleic acid molecules that encode the protein having SEQ ID NO: 2, 4, 6, 8 or 10 and the related proteins herein described, preferably in isolated form. As used herein, "nucleic acid" is defined as RNA or DNA that encodes a protein or peptide as defined above, is complementary to a nucleic acid

10

15

20

25

30

sequence encoding such peptides; hybridizes to the nucleic acid of SEQ ID NO: 1, 3, 5, 7 or 9 and remains stably bound to it under appropriate stringency conditions; encodes a polypeptide sharing at least about 50%, 60%, 70% or 75%, preferably at least about 80%, more preferably at least about 85%, and most preferably at least about 90%, 95%, 98%, 99%, 99.5% or more identity with the peptide sequence of SEQ ID NO: 2 or 4, or a polypeptide sharing at least about 50%, 60%, 70% or 75%, preferably at least about 80-90%, more preferably at least about 91-92%, and most preferably at least about 93%, 95%, 98%, 99% or more identity with the peptide sequence of SEQ ID NO: 6 or 8; or exhibits at least 50%, 60%, 70% or 75%, preferably at least about 80%, more preferably at least about 85%, and even more preferably at least about 90%, 95%, 98%, 99%, 99.5% or more nucleotide sequence identity over the open reading frames of SEQ ID NO: 1 or 3, or at least 50%, 60%, 70% or 75%, preferably at least about 80-90%, more preferably at least about 91-92%, and even more preferably at least about 93%, 95%, 98%, 99% or more nucleotide sequence identity over the open reading frames of SEQ ID NO: 5, 7 or 9.

The present invention further includes isolated nucleic acid molecules that specifically hybridize to the complement of SEQ ID NO: 1, 3, 5, 7 or 9, particularly molecules that specifically hybridize over the open reading frames. Such molecules that specifically hybridize to the complement of SEQ ID NO: 1, 3, 5, 7 or 9 typically do so under stringent hybridization conditions.

Specifically contemplated are genomic DNA, cDNA, mRNA and antisense molecules, as well as nucleic acids based on alternative backbones or including alternative bases, whether derived from natural sources or synthesized. Such hybridizing or complementary nucleic acids, however, are defined further as being novel and unobvious over any prior art nucleic acid including that which encodes, hybridizes under appropriate stringency conditions, or is complementary to nucleic acid encoding a protein according to the present invention.

Homology or identity at the nucleotide or amino acid sequence level is determined by BLAST (Basic Local Alignment Search Tool) analysis using the algorithm employed by the programs blastp, blastn, blastx, tblastn and tblastx (Altschul et al., (1997) Nucleic Acids Res 25:3389-3402, and Karlin et al., (1990) Proc

10

15

20

25

30

Natl Acad Sci USA 87:2264-2268, both fully incorporated by reference) which are tailored for sequence similarity searching. The approach used by the BLAST program is to first consider similar segments, with and without gaps, between a query sequence and a database sequence, then to evaluate the statistical significance of all matches that are identified and finally to summarize only those matches which satisfy a preselected threshold of significance. For a discussion of basic issues in similarity searching of sequence databases, see Altschul et al., (1994) Nature Genetics 6: 119-129 which is fully incorporated by reference. The search parameters for histogram, descriptions, alignments, expect (i.e., the statistical significance threshold for reporting matches against database sequences), cutoff, matrix and filter (low complexity) are at the default settings. The default scoring matrix used by blastp, blastx, tblastn, and tblastx is the BLOSUM62 matrix (Henikoff et al., (1992) Proc Natl Acad Sci USA 89:10915-10919, fully incorporated by reference), recommended for query sequences over 85 nucleotides or amino acids in length.

For blastn, the scoring matrix is set by the ratios of M (i.e., the reward score for a pair of matching residues) to N (i.e., the penalty score for mismatching residues), wherein the default values for M and N are 5 and -4, respectively. Four blastn parameters were adjusted as follows: Q=10 (gap creation penalty); R=10 (gap extension penalty); wink=1 (generates word hits at every wink<sup>th</sup> position along the query); and gapw=16 (sets the window width within which gapped alignments are generated). The equivalent Blastp parameter settings were Q=9; R=2; wink=1; and gapw=32. A Bestfit comparison between sequences, available in the GCG package version 10.0, uses DNA parameters GAP=50 (gap creation penalty) and LEN=3 (gap extension penalty) and the equivalent settings in protein comparisons are GAP=8 and LEN=2.

"Stringent conditions" are those that (1) employ low ionic strength and high temperature for washing, for example, 0.015 M NaCl/0.0015 M sodium citrate/0.1% SDS at 50 °C, or (2) employ during hybridization a denaturing agent such as formamide, for example, 50% (vol/vol) formamide with 0.1% bovine serum albumin/0.1% Ficoll/0.1% polyvinylpyrrolidone/50 mM sodium phosphate buffer at pH 6.5 with 750 mM NaCl, 75 mM sodium citrate at 42 °C. Another example is hybridization in 50%

10

15

20

25

30

formamide, 5× SSC (0.75 M NaCl, 0.075 M sodium citrate), 50 mM sodium phosphate (pH 6.8), 0.1% sodium pyrophosphate, 5× Denhardt's solution, sonicated salmon sperm DNA (50 μg/ml), 0.1% SDS, and 10% dextran sulfate at 42 °C, with washes at 42 °C in 0.2× SSC and 0.1% SDS. A skilled artisan can readily determine and vary the stringency conditions appropriately to obtain a clear and detectable hybridization signal. Preferred molecules are those that hybridize under the above conditions to the complement of SEQ ID NO: 1, 3, 5 or 7 and which encode a functional or full-length protein. Even more preferred hybridizing molecules are those that hybridize under the above conditions to the complement strand of the open reading frame of SEQ ID NO: 1, 3, 5, 7 or 9.

As used herein, a nucleic acid molecule is said to be "isolated" when the nucleic acid molecule is substantially separated from contaminant nucleic acid molecules encoding other polypeptides.

The present invention further provides fragments of the disclosed nucleic acid molecules. As used herein, a fragment of a nucleic acid molecule refers to a small portion of the coding or non-coding sequence. The size of the fragment will be determined by the intended use. For example, if the fragment is chosen so as to encode an active portion of the protein, the fragment will need to be large enough to encode the functional region(s) of the protein. For instance, fragments which encode peptides corresponding to predicted antigenic regions may be prepared. If the fragment is to be used as a nucleic acid probe or PCR primer, then the fragment length is chosen so as to obtain a relatively small number of false positives during probing/priming (see the discussion in Section H).

Fragments of the nucleic acid molecules of the present invention (i.e., synthetic oligonucleotides) that are used as probes or specific primers for the polymerase chain reaction (PCR), or to synthesize gene sequences encoding proteins of the invention, can easily be synthesized by chemical techniques, for example, the phosphoramidite method of Matteucci et al., ((1981) J Am Chem Soc 103:3185-3191) or using automated synthesis methods. In addition, larger DNA segments can readily be prepared by well known methods, such as synthesis of a group of oligonucleotides that define various

10

15

20

25

30

modular segments of the gene, followed by ligation of oligonucleotides to build the complete modified gene.

The nucleic acid molecules of the present invention may further be modified so as to contain a detectable label for diagnostic and probe purposes. A variety of such labels are known in the art and can readily be employed with the encoding molecules herein described. Suitable labels include, but are not limited to, biotin, radiolabeled or fluorescently labeled nucleotides and the like. A skilled artisan can readily employ any such label to obtain labeled variants of the nucleic acid molecules of the invention.

#### C. Isolation of Other Related Nucleic Acid Molecules

As described above, the identification and characterization of the nucleic acid molecule having SEQ ID NO: 1, 3, 5, 7 or 9 allows a skilled artisan to isolate nucleic acid molecules that encode other members of the protein family in addition to the sequences herein described. Further, the presently disclosed nucleic acid molecules allow a skilled artisan to isolate nucleic acid molecules that encode other members of the family of proteins in addition to the proteins having SEQ ID NO: 2, 4, 6, 8 or 10.

For instance, a skilled artisan can readily use the amino acid sequence of SEQ ID NO: 2, 4, 6, 8 or 10 to generate antibody probes to screen expression libraries prepared from appropriate cells. Typically, polyclonal antiserum from mammals such as rabbits immunized with the purified protein (as described below) or monoclonal antibodies can be used to probe a mammalian cDNA or genomic expression library, such as lambda gtll library, to obtain the appropriate coding sequence for other members of the protein family. The cloned cDNA sequence can be expressed as a fusion protein, expressed directly using its own control sequences, or expressed by constructions using control sequences appropriate to the particular host used for expression of the enzyme.

Alternatively, a portion of the coding sequence herein described can be synthesized and used as a probe to retrieve DNA encoding a member of the protein family from any mammalian organism. Oligomers containing approximately 18-20 nucleotides (encoding about a 6-7 amino acid stretch) are prepared and used to screen genomic DNA or cDNA libraries to obtain hybridization under stringent conditions or

10

15

20

25

30

conditions of sufficient stringency to eliminate an undue level of false positives.

Additionally, pairs of oligonucleotide primers can be prepared for use in a polymerase chain reaction (PCR) to selectively clone an encoding nucleic acid molecule. A PCR denature/anneal/extend cycle for using such PCR primers is well known in the art and can readily be adapted for use in isolating other encoding nucleic acid molecules.

Nucleic acid molecules encoding other members of the protein family may also be identified in existing genomic or other sequence information using any available computational method, including but not limited to: PSI-BLAST (Altschul et al., (1997) Nucleic Acids Res 25:3389-3402); PHI-BLAST (Zhang et al., (1998) Nucleic Acids Res 26:3986-3990), 3D-PSSM (Kelly et al., (2000) J Mol Biol 299(2):499-520); and other computational analysis methods (Shi et al., (1999) Biochem Biophys Res Commun 262(1):132-138 and Matsunami et. al., (2000) Nature 404(6778):601-604.

#### D. rDNA molecules Containing a Nucleic Acid Molecule

The present invention further provides recombinant DNA molecules (rDNAs) that contain a coding sequence. As used herein, a rDNA molecule is a DNA molecule that has been subjected to molecular manipulation in situ. Methods for generating rDNA molecules are well known in the art, for example, see Sambrook et al., Molecular Cloning - A Laboratory Manual, Third Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001. In the preferred rDNA molecules, a coding DNA sequence is operably linked to expression control sequences and/or vector sequences.

The choice of vector and/or expression control sequences to which one of the protein family encoding sequences of the present invention is operably linked depends directly, as is well known in the art, on the functional properties desired, e.g., protein expression, and the host cell to be transformed. A vector contemplated by the present invention is at least capable of directing the replication or insertion into the host chromosome, and preferably also expression, of the structural gene included in the rDNA molecule.

Expression control elements that are used for regulating the expression of an operably linked protein encoding sequence are known in the art and include, but are not

10

15

20

25

30

limited to, inducible promoters, constitutive promoters, secretion signals, and other regulatory elements. Preferably, the inducible promoter is readily controlled, such as being responsive to a nutrient in the host cell's medium.

In one embodiment, the vector containing a coding nucleic acid molecule will include a prokaryotic replicon, *i.e.*, a DNA sequence having the ability to direct autonomous replication and maintenance of the recombinant DNA molecule extrachromosomally in a prokaryotic host cell, such as a bacterial host cell, transformed therewith. Such replicons are well known in the art. In addition, vectors that include a prokaryotic replicon may also include a gene whose expression confers a detectable marker such as a drug resistance. Typical bacterial drug resistance genes are those that confer resistance to ampicillin, kanamycin, chloramphenicol or tetracycline.

Vectors that include a prokaryotic replicon can further include a prokaryotic or bacteriophage promoter capable of directing the expression (transcription and translation) of the coding gene sequences in a bacterial host cell, such as *E. coli*. A promoter is an expression control element formed by a DNA sequence that permits binding of RNA polymerase and transcription to occur. Promoter sequences compatible with bacterial hosts are typically provided in plasmid vectors containing convenient restriction sites for insertion of a DNA segment of the present invention. Typical of such vector plasmids are pUC8, pUC9, pBR322 and pBR329 available from BioRad Laboratories, (Richmond, CA), pPL and pKK223 available from Pharmacia (Piscataway, NJ).

Expression vectors compatible with eukaryotic cells, preferably those compatible with vertebrate cells, such as liver cells, can also be used to form rDNA molecules that contain a coding sequence. Eukaryotic cell expression vectors, including viral vectors, are well known in the art and are available from several commercial sources. Typically, such vectors are provided containing convenient restriction sites for insertion of the desired DNA segment. Typical of such vectors are pSVL and pKSV-10 (Pharmacia), pBPV-1/pML2d (International Biotechnologies, Inc.), pTDT1 (ATCC, #31255), the vector pCDM8 described herein, and the like eukaryotic expression vectors. Vectors may be modified to include liver cell specific promoters if

needed.

Eukaryotic cell expression vectors used to construct the rDNA molecules of the present invention may further include a selectable marker that is effective in an eukaryotic cell, preferably a drug resistance selection marker. A preferred drug resistance marker is the gene whose expression results in neomycin resistance, *i.e.*, the neomycin phosphotransferase (neo) gene. (Southern et al., (1982) J Mol Anal Genet 1:327-341) Alternatively, the selectable marker can be present on a separate plasmid, and the two vectors are introduced by co-transfection of the host cell, and selected by culturing in the appropriate drug for the selectable marker.

10

15

20

25

30

5

#### E. Host Cells Containing an Exogenously Supplied Coding Nucleic Acid Molecule

The present invention further provides host cells transformed with a nucleic acid molecule that encodes a protein of the present invention. The host cell can be either prokaryotic or eukaryotic. Eukaryotic cells useful for expression of a protein of the invention are not limited, so long as the cell line is compatible with cell culture methods and compatible with the propagation of the expression vector and expression of the gene product. Preferred eukaryotic host cells include, but are not limited to, yeast, insect and mammalian cells, preferably vertebrate cells such as those from a mouse, rat, monkey or human cell line. Preferred eukaryotic host cells include Chinese hamster ovary (CHO) cells available from the ATCC as CCL61, NIH Swiss mouse embryo cells (NIH/3T3) available from the ATCC as CRL 1658, baby hamster kidney cells (BHK), and the like eukaryotic tissue culture cell lines.

Any prokaryotic host can be used to express a rDNA molecule encoding a protein of the invention. The preferred prokaryotic host is *E. coli*.

Transformation of appropriate cell hosts with a rDNA molecule of the present invention is accomplished by well known methods that typically depend on the type of vector used and host system employed. With regard to transformation of prokaryotic host cells, electroporation and salt treatment methods are typically employed (see, for example, Cohen et al., (1972) Proc Natl Acad Sci USA 69:2110; and Sambrook et al., supra). With regard to transformation of vertebrate cells with vectors containing

10

15

20

25

30

rDNAs, electroporation, cationic lipid or salt treatment methods are typically employed, see, for example, Graham et al., (1973) Virol 52:456; Wigler et al., (1979) Proc Natl Acad Sci USA 76;1373-1376.

Successfully transformed cells, i.e., cells that contain a rDNA molecule of the present invention, can be identified by well known techniques including the selection for a selectable marker. For example, cells resulting from the introduction of an rDNA of the present invention can be cloned to produce single colonies. Cells from those colonies can be harvested, lysed and their DNA content examined for the presence of the rDNA using a method such as that described by Southern, (1975) J Mol Biol 98:503 or Berent et al., (1985) Biotech 3:208, or the proteins produced from the cell assayed via an immunological method.

#### F. Production of Recombinant Proteins using a rDNA Molecule

The present invention further provides methods for producing a protein of the invention using nucleic acid molecules herein described. In general terms, the production of a recombinant form of a protein typically involves the following steps:

First, a nucleic acid molecule is obtained that encodes a protein of the invention, such as a nucleic acid molecule comprising, consisting essentially of or consisting of SEQ ID NO: 1 or SEQ ID NO: 3; nucleotides 155-421 or 155-418 of SEQ ID NO: 1; nucleotides 139-405 or 139-402 of SEQ ID NO: 3; SEQ ID NO: 5, SEQ ID NO: 7 or SEQ ID NO: 9; nucleotides 32-1387 or 32-1384 of SEQ ID NO: 5; nucleotides 41-1504 or 41-1501 of SEQ ID NO: 7; or nucleotides 31-1554 or 31-1551 of SEQ ID NO: 9. If the encoding sequence is uninterrupted by introns, as are these open-reading-frames, it is directly suitable for expression in any host.

The nucleic acid molecule is then preferably placed in operable linkage with suitable control sequences, as described above, to form an expression unit containing the protein open reading frame. The expression unit is used to transform a suitable host and the transformed host is cultured under conditions that allow the production of the recombinant protein. Optionally the recombinant protein is isolated from the medium or from the cells; recovery and purification of the protein may not be necessary in some

10

15

20

25

30

instances where some impurities may be tolerated.

Each of the foregoing steps can be done in a variety of ways. For example, the desired coding sequences may be obtained from genomic fragments and used directly in appropriate hosts. The construction of expression vectors that are operable in a variety of hosts is accomplished using appropriate replicons and control sequences, as set forth above. The control sequences, expression vectors, and transformation methods are dependent on the type of host cell used to express the gene and were discussed in detail earlier. Suitable restriction sites can, if not normally available, be added to the ends of the coding sequence so as to provide an excisable gene to insert into these vectors. A skilled artisan can readily adapt any host/expression system known in the art for use with the nucleic acid molecules of the invention to produce recombinant protein.

#### G. Methods to Identify Binding Partners

Another embodiment of the present invention provides methods for isolating and identifying binding partners of proteins of the invention. In general, a protein of the invention is mixed with a potential binding partner or an extract or fraction of a cell under conditions that allow the association of potential binding partners with the protein of the invention. After mixing, peptides, polypeptides, proteins or other molecules that have become associated with a protein of the invention are separated from the mixture. The binding partner that bound to the protein of the invention can then be removed and further analyzed. To identify and isolate a binding partner, the entire protein, for instance a protein comprising the entire amino acid sequence of SEQ ID NO: 2, 4, 6, 8 or 10 can be used. Alternatively, a fragment of the protein can be used.

As used herein, a cellular extract refers to a preparation or fraction which is made from a lysed or disrupted cell. The preferred source of cellular extracts will be cells derived from human liver tumors or transformed liver cells, for instance, biopsy tissue or tissue culture cells from hepatic carcinomas. Alternatively, cellular extracts may be prepared from normal tissue or available cell lines, particularly liver-derived cell lines.

A variety of methods can be used to obtain an extract of a cell. Cells can be

10

15

20

25

30

disrupted using either physical or chemical disruption methods. Examples of physical disruption methods include, but are not limited to, sonication and mechanical shearing. Examples of chemical lysis methods include, but are not limited to, detergent lysis and enzyme lysis. A skilled artisan can readily adapt methods for preparing cellular extracts in order to obtain extracts for use in the present methods.

Once an extract of a cell is prepared, the extract is mixed with the protein of the invention under conditions in which association of the protein with the binding partner can occur. A variety of conditions can be used, the most preferred being conditions that closely resemble conditions found in the cytoplasm of a human cell. Features such as osmolarity, pH, temperature, and the concentration of cellular extract used, can be varied to optimize the association of the protein with the binding partner.

After mixing under appropriate conditions, the bound complex is separated from the mixture. A variety of techniques can be utilized to separate the mixture. For example, antibodies specific to a protein of the invention can be used to immunoprecipitate the binding partner complex. Alternatively, standard chemical separation techniques such as chromatography and density/sediment centrifugation can be used.

After removal of non-associated cellular constituents found in the extract, the binding partner can be dissociated from the complex using conventional methods. For example, dissociation can be accomplished by altering the salt concentration or pH of the mixture.

To aid in separating associated binding partner pairs from the mixed extract, the protein of the invention can be immobilized on a solid support. For example, the protein can be attached to a nitrocellulose matrix or acrylic beads. Attachment of the protein to a solid support aids in separating peptide/binding partner pairs from other constituents found in the extract. The identified binding partners can be either a single protein or a complex made up of two or more proteins. Alternatively, binding partners may be identified using a Far-Western assay according to the procedures of Takayama et al., (1997) Methods Mol Biol 69:171-184 or Sauder et al., (1996) J Gen Virol 77:991-996 or identified through the use of epitope tagged proteins or GST fusion proteins.

Alternatively, the nucleic acid molecules of the invention can be used in a yeast two-hybrid system or other in vivo protein-protein detection system. The yeast twohybrid system has been used to identify other protein partner pairs and can readily be adapted to employ the nucleic acid molecules herein described.

5

10

15

20

25

30

#### H. Methods to Identify Agents that Modulate the Expression a Nucleic Acid **Encoding the Genes Associated with Liver Cancer**

Another embodiment of the present invention provides methods for identifying agents that modulate the expression of a nucleic acid encoding a protein of the invention such as a protein having the amino acid sequence of SEQ ID NO: 2, 4, 6, 8 or 10. Such assays may utilize any available means of monitoring for changes in the expression level of the nucleic acids of the invention. As used herein, an agent is said to modulate the expression of a nucleic acid of the invention if it is capable of up- or down-regulating expression of the nucleic acid in a cell.

In one assay format, cell lines that contain reporter gene fusions between nucleotides from within the open reading frame defined by nucleotides 155-421 of SEO ID NO: 1, nucleotides 139-405 of SEQ ID NO: 3, nucleotides 32-1387 of SEQ ID NO: 5, nucleotides 41-1504 of SEQ ID NO: 7, or nucleotides 31-1554 of SEQ ID NO: 9 and/or the 5'and/or 3' regulatory elements and any assayable fusion partner may be prepared. Numerous assayable fusion partners are known and readily available including the firefly luciferase gene and the gene encoding chloramphenicol acetyltransferase (Alam et al., (1990) Anal Biochem 188:245-254). Cell lines containing the reporter gene fusions are then exposed to the agent to be tested under appropriate conditions and time. Differential expression of the reporter gene between samples exposed to the agent and control samples identifies agents which modulate the expression of a nucleic acid of the invention.

Additional assay formats may be used to monitor the ability of the agent to modulate the expression of a nucleic acid encoding a protein of the invention, such as the protein having SEQ ID NO: 2, 4, 6, 8 or 10. For instance, mRNA expression may be monitored directly by hybridization to the nucleic acids of the invention. Cell lines are

10

15

20

25

30

exposed to the agent to be tested under appropriate conditions and time and total RNA or mRNA is isolated by standard procedures such those disclosed in Sambrook *et al.*, Molecular Cloning - A Laboratory Manual, Third Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001.

The preferred cells will be those derived from human liver tissue, for instance, liver biopsy tissue or cultured cells from patients with liver cancer, or liver cancer and cirrhosis, or liver cancer and hepatitis. Cell lines such as ATCC hepatocellular carcinoma cell lines Catalogue Nos. HB-8064, HB-8065 or CRL-10741 may be used. Alternatively, other available cells or cell lines may be used.

Probes to detect differences in RNA expression levels between cells exposed to the agent and control cells may be prepared from the nucleic acids of the invention. It is preferable, but not necessary, to design probes which hybridize only with target nucleic acids under conditions of high stringency. Only highly complementary nucleic acid hybrids form under conditions of high stringency. Accordingly, the stringency of the assay conditions determines the amount of complementarity which should exist between two nucleic acid strands in order to form a hybrid. Stringency should be chosen to maximize the difference in stability between the probe:target hybrid and probe:non-target hybrids.

Probes may be designed from the nucleic acids of the invention through methods known in the art. For instance, the G+C content of the probe and the probe length can affect probe binding to its target sequence. Methods to optimize probe specificity are commonly available in Sambrook et al., supra, or Ausubel et al., Short Protocols in Molecular Biology, Fourth Ed., John Wiley & Sons, Inc., New York, 1999.

Hybridization conditions are modified using known methods, such as those described by Sambrook et al. and Ausubel et al. as required for each probe. Hybridization of total cellular RNA or RNA enriched for polyA RNA can be accomplished in any available format. For instance, total cellular RNA or RNA enriched for polyA RNA can be affixed to a solid support and the solid support exposed to at least one probe comprising at least one, or part of one of the sequences of the invention under conditions in which the probe will specifically hybridize. Alternatively,

10

15

20

25

30

nucleic acid fragments comprising at least one, or part of one of the sequences of the invention can be affixed to a solid support, such as a silicon chip, porous glass wafer or membrane. The solid support can then be exposed to total cellular RNA or polyA RNA from a sample under conditions in which the affixed sequences will specifically hybridize. Such solid supports and hybridization methods are widely available, for example, those disclosed by Beattie, (1995) WO 95/11755. By examining for the ability of a given probe to specifically hybridize to an RNA sample from an untreated cell population and from a cell population exposed to the agent, agents which up- or down-regulate the expression of a nucleic acid encoding the protein having the sequence of SEQ ID NO: 2, 4, 6, 8 or 10 are identified.

Hybridization for qualitative and quantitative analysis of mRNAs may also be carried out by using a RNase Protection Assay (*i.e.*, RPA, see Ma *et al.*, (1996) *Methods* 10:273-238). Briefly, an expression vehicle comprising cDNA encoding the gene product and a phage specific DNA dependent RNA polymerase promoter (*e.g.*, T7, T3 or SP6 RNA polymerase) is linearized at the 3' end of the cDNA molecule, downstream from the phage promoter, wherein such a linearized molecule is subsequently used as a template for synthesis of a labeled antisense transcript of the cDNA by *in vitro* transcription. The labeled transcript is then hybridized to a mixture of isolated RNA (*i.e.*, total or fractionated mRNA) by incubation at 45 °C overnight in a buffer comprising 80% formamide, 40 mM Pipes, pH 6.4, 0.4 M NaCl and 1 mM EDTA. The resulting hybrids are then digested in a buffer comprising 40 μg/ml ribonuclease A and 2 μg/ml ribonuclease. After deactivation and extraction of extraneous proteins, the samples are loaded onto urea/polyacrylamide gels for analysis.

In another assay, to identify agents which affect the expression of the instant gene products, cells or cell lines are first identified which express the gene products of the invention physiologically. Cell and/or cell lines so identified would be expected to comprise the necessary cellular machinery such that the fidelity of modulation of the transcriptional apparatus is maintained with regard to exogenous contact of agent with appropriate surface transduction mechanisms and/or the cytosolic cascades. Further, such cells or cell lines would be transduced or transfected with an expression vehicle

10

15

20

25

30

(e.g., a plasmid or viral vector) construct comprising an operable non-translated 5'promoter-containing end of the structural gene encoding the instant gene products fused to one or more antigenic fragments, which are peculiar to the instant gene products, wherein said fragments are under the transcriptional control of said promoter and are expressed as polypeptides whose molecular weight can be distinguished from the naturally occurring polypeptides or may further comprise an immunologically distinct tag or other detectable marker. Such a process is well known in the art (see Sambrook et al., supra).

Cells or cell lines transduced or transfected as outlined above are then contacted with agents under appropriate conditions. For example, the agent in a pharmaceutically acceptable excipient is contacted with cells in an aqueous physiological buffer such as phosphate buffered saline (PBS) at physiological pH, Eagles balanced salt solution (BSS) at physiological pH, PBS or BSS comprising serum or conditioned media comprising PBS or BSS and/or serum incubated at 37 °C. Said conditions may be modulated as deemed necessary by one of skill in the art. Subsequent to contacting the cells with the agent, said cells will be disrupted and the polypeptides of the lysate are fractionated such that a polypeptide fraction is pooled and contacted with an antibody to be further processed by immunological assay (e.g., ELISA, immunoprecipitation or Western blot). The pool of proteins isolated from the "agent-contacted" sample will be compared with a control sample where only the excipient is contacted with the cells and an increase or decrease in the immunologically generated signal from the "agent-contacted" sample compared to the control will be used to distinguish the effectiveness of the agent.

# H. Methods to Identify Agents that Modulate the Level or at Least One Activity of the Liver Cancer Associated Proteins

Another embodiment of the present invention provides methods for identifying agents that modulate the level or at least one activity of a protein of the invention such as the protein having the amino acid sequence of SEQ ID NO: 2, 4, 6, 8 or 10. Such methods or assays may utilize any means of monitoring or detecting the desired activity.

In one format, the relative amounts of a protein of the invention between a cell

10

15

20

25

30

population that has been exposed to the agent to be tested compared to an un-exposed control cell population may be assayed. In this format, probes such as specific antibodies are used to monitor the differential expression of the protein in the different cell populations. Cell lines or populations are exposed to the agent to be tested under appropriate conditions and time. Cellular lysates may be prepared from the exposed cell line or population and a control, unexposed cell line or population. The cellular lysates are then analyzed with the probe.

Antibody probes are prepared by immunizing suitable mammalian hosts in appropriate immunization protocols using the peptides, polypeptides or proteins of the invention if they are of sufficient length, or, if desired, or if required to enhance immunogenicity, conjugated to suitable carriers. Methods for preparing immunogenic conjugates with carriers such as BSA, KLH, or other carrier proteins are well known in the art. In some circumstances, direct conjugation using, for example, carbodiimide reagents may be effective; in other instances linking reagents such as those supplied by Pierce Chemical Co. (Rockford, IL), may be desirable to provide accessibility to the hapten. The hapten peptides can be extended at either the amino or carboxy terminus with a cysteine residue or interspersed with cysteine residues, for example, to facilitate linking to a carrier. Administration of the immunogens is conducted generally by injection over a suitable time period and with use of suitable adjuvants, as is generally understood in the art. During the immunization schedule, titers of antibodies are taken to determine adequacy of antibody formation.

While the polyclonal antisera produced in this way may be satisfactory for some applications, for pharmaceutical compositions, use of monoclonal preparations is preferred. Immortalized cell lines which secrete the desired monoclonal antibodies may be prepared using the standard method of Kohler and Milstein ((1975) Nature 256:495-497) or modifications which effect immortalization of lymphocytes or spleen cells, as is generally known. The immortalized cell lines secreting the desired antibodies are screened by immunoassay in which the antigen is the peptide hapten, polypeptide or protein. When the appropriate immortalized cell culture secreting the desired antibody is identified, the cells can be cultured either in vitro or by production in

ascites fluid.

5

10

15

20

25

30

The desired monoclonal antibodies are then recovered from the culture supernatant or from the ascites supernatant. Fragments of the monoclonal antibodies or the polyclonal antisera which contain the immunologically significant (antigen-binding) portion can be used as antagonists, as well as the intact antibodies. Use of immunologically reactive (antigen-binding) antibody fragments, such as the Fab, Fab', or  $F(ab')_2$  fragments is often preferable, especially in a therapeutic context, as these fragments are generally less immunogenic than the whole immunoglobulin.

The antibodies or antigen-binding fragments may also be produced, using current technology, by recombinant means. Antibody regions that bind specifically to the desired regions of the protein can also be produced in the context of chimeras with multiple species origin, such as humanized antibodies.

Agents that are assayed in the above method can be randomly selected or rationally selected or designed. As used herein, an agent is said to be randomly selected when the agent is chosen randomly without considering the specific sequences involved in the association of a protein of the invention alone or with its associated substrates, binding partners, etc. An example of randomly selected agents is the use a chemical library or a peptide combinatorial library, or a growth broth of an organism.

As used herein, an agent is said to be rationally selected or designed when the agent is chosen on a nonrandom basis which takes into account the sequence of the target site and/or its conformation in connection with the agent's action. Agents can be rationally selected or rationally designed by utilizing the peptide sequences that make up these sites. For example, a rationally selected peptide agent can be a peptide whose amino acid sequence is identical to or a derivative of any functional consensus site.

The agents of the present invention can be, as examples, peptides, small molecules, vitamin derivatives, as well as carbohydrates. Dominant negative proteins, DNAs encoding these proteins, antibodies to these proteins, peptide fragments of these proteins or mimics of these proteins may be introduced into cells to affect function. "Mimic" used herein refers to the modification of a region or several regions of a peptide molecule to provide a structure chemically different from the parent peptide but

biology and Biotechnology, Meyers, ed., pp. 659-664, VCH Publishers, Inc., New York, 1995). A skilled artisan can readily recognize that there is no limit as to the structural nature of the agents of the present invention.

The peptide agents of the invention can be prepared using standard solid phase (or solution phase) peptide synthesis methods, as is known in the art. In addition, the DNA encoding these peptides may be synthesized using commercially available oligonucleotide synthesis instrumentation and produced recombinantly using standard recombinant production systems. The production using solid phase peptide synthesis is necessitated if non-gene-encoded amino acids are to be included.

Another class of agents of the present invention are antibodies immunoreactive with critical positions of proteins of the invention. Antibody agents are obtained by immunization of suitable mammalian subjects with peptides, containing as antigenic regions, those portions of the protein intended to be targeted by the antibodies.

15

20

25

30

10

5

# J. Uses for Agents that Modulate the Expression or at Least one Activity of the Proteins Associated with Liver Cancer

As provided in the Examples, the proteins and nucleic acids of the invention, such as the proteins having the amino acid sequence of SEQ ID NO: 2, 4, 6, 8 or 10, are differentially expressed in cancerous liver tissue. Agents that up- or down- regulate or modulate the expression of the protein or at least one activity of the protein, such as agonists or antagonists, of may be used to modulate biological and pathologic processes associated with the protein's function and activity.

As used herein, a subject can be any mammal, so long as the mammal is in need of modulation of a pathological or biological process mediated by a protein of the invention. The term "mammal" is defined as an individual belonging to the class Mammalia. The invention is particularly useful in the treatment of human subjects.

Pathological processes refer to a category of biological processes which produce a deleterious effect. For example, expression of a protein of the invention may be associated with liver cell growth or hyperplasia. As used herein, an agent is said to

10

15

20

25

30

modulate a pathological process when the agent reduces the degree or severity of the process. For instance, liver cancer may be prevented or disease progression modulated by the administration of agents which up- or down-regulate or modulate in some way the expression or at least one activity of a protein of the invention.

The agents of the present invention can be provided alone, or in combination with other agents that modulate a particular pathological process. For example, an agent of the present invention can be administered in combination with other known drugs. As used herein, two agents are said to be administered in combination when the two agents are administered simultaneously or are administered independently in a fashion such that the agents will act at the same time.

The agents of the present invention can be administered via parenteral, subcutaneous, intravenous, intramuscular, intraperitoneal, transdermal, or buccal routes. Alternatively, or concurrently, administration may be by the oral route. The dosage administered will be dependent upon the age, health, and weight of the recipient, kind of concurrent treatment, if any, frequency of treatment, and the nature of the effect desired.

The present invention further provides compositions containing one or more agents which modulate expression or at least one activity of a protein of the invention. While individual needs vary, determination of optimal ranges of effective amounts of each component is within the skill of the art. Typical dosages comprise 0.1 to 100 µg/kg body wt. The preferred dosages comprise 0.1 to 10 µg/kg body wt. The most preferred dosages comprise 0.1 to 1 µg/kg body wt.

In addition to the pharmacologically active agent, the compositions of the present invention may contain suitable pharmaceutically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically for delivery to the site of action. Suitable formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form, for example, water-soluble salts. In addition, suspensions of the active compounds as appropriate oily injection suspensions may be administered. Suitable lipophilic solvents or vehicles include fatty oils, for example, sesame oil, or synthetic fatty acid esters, for example, ethyl oleate or triglycerides.

10

15

25

30

Aqueous injection suspensions may contain substances which increase the viscosity of the suspension include, for example, sodium carboxymethyl cellulose, sorbitol, and/or dextran. Optionally, the suspension may also contain stabilizers. Liposomes can also be used to encapsulate the agent for delivery into the cell.

The pharmaceutical formulation for systemic administration according to the invention may be formulated for enteral, parenteral or topical administration. Indeed, all three types of formulations may be used simultaneously to achieve systemic administration of the active ingredient.

Suitable formulations for oral administration include hard or soft gelatin capsules, pills, tablets, including coated tablets, elixirs, suspensions, syrups or inhalations and controlled release forms thereof.

In practicing the methods of this invention, the compounds of this invention may be used alone or in combination, or in combination with other therapeutic or diagnostic agents. In certain preferred embodiments, the compounds of this invention may be coadministered along with other compounds typically prescribed for these conditions according to generally accepted medical practice. The compounds of this invention can be utilized *in vivo*, ordinarily in mammals, such as humans, sheep, horses, cattle, pigs, dogs, cats, rats and mice, or *in vitro*.

#### 20 K. Transgenic Animals

Transgenic animals containing mutant, knock-out or modified genes corresponding to the cDNA sequence of SEQ ID NO: 1, 3, 5, 7 or 9, or the open reading frame encoding the polypeptide sequence of SEQ ID NO: 2, 4, 6, 8 or 10 or fragments thereof having a consecutive sequence of at least about 3, 4, 5, 6, 10, 15, 20, 25, 30, 35 or more amino acid residues, are also included in the invention. Transgenic animals are genetically modified animals into which recombinant, exogenous or cloned genetic material has been experimentally transferred. Such genetic material is often referred to as a "transgene." The nucleic acid sequence of the transgene, in this case a form of SEQ ID NO: 1, 3, 5, 7 or 9, may be integrated either at a locus of a genome where that particular nucleic acid sequence is not otherwise normally found or at the normal locus

10

15

20

25

30

for the transgene. The transgene may consist of nucleic acid sequences derived from the genome of the same species or of a different species than the species of the target animal.

In some embodiments, transgenic animals in which all or a portion of a gene comprising SEQ ID NO: 1, 3, 5, 7 or 9 is deleted may be constructed. In those cases where the gene corresponding to SEQ ID NO: 1, 3, 5, 7 or 9 contains one or more introns, the entire gene- all exons, introns and the regulatory sequences- may be deleted. Alternatively, less than the entire gene may be deleted. For example, a single exon and/or intron may be deleted, so as to create an animal expressing a modified version of a protein of the invention.

The term "germ cell line transgenic animal" refers to a transgenic animal in which the genetic alteration or genetic information was introduced into a germ line cell, thereby conferring the ability of the transgenic animal to transfer the genetic information to offspring. If such offspring in fact possess some or all of that alteration or genetic information, then they too are transgenic animals.

The alteration or genetic information may be foreign to the species of animal to which the recipient belongs, foreign only to the particular individual recipient, or may be genetic information already possessed by the recipient. In the last case, the altered or introduced gene may be expressed differently than the native gene.

Transgenic animals can be produced by a variety of different methods including transfection, electroporation, microinjection, gene targeting in embryonic stem cells and recombinant viral and retroviral infection (see, e.g., U.S. Patent No. 4,736,866; U.S. Patent No. 5,602,307; Mullins et al., (1993) Hypertension 22:630-633; Brenin et al., (1997) Surg Oncol 6:99-110; Recombinant Gene Expression Protocols (Methods in Molecular Biology, Vol. 62), Tuan, ed., Humana Press, Totowa, NJ, 1997).

A number of recombinant or transgenic mice have been produced, including those which express an activated oncogene sequence (U.S. Patent No. 4,736,866); express simian SV40 T-antigen (U.S. Patent No. 5,728,915); lack the expression of interferon regulatory factor 1 (IRF-1) (U.S. Patent No. 5,731,490); exhibit dopaminergic dysfunction (U.S. Patent No. 5,723,719); express at least one human gene which participates in blood pressure control (U.S. Patent No. 5,731,489); display greater

10

15

20

25

30

similarity to the conditions existing in naturally occurring Alzheimer's disease (U.S. Patent No. 5,720,936); have a reduced capacity to mediate cellular adhesion (U.S. Patent No. 5,602,307); possess a bovine growth hormone gene (Clutter et al., (1996) Genetics 143:1753-1760); or, are capable of generating a fully human antibody response (McCarthy (1997) Lancet 349:405).

While mice and rats remain the animals of choice for most transgenic experimentation, in some instances it is preferable or even necessary to use alternative animal species. Transgenic procedures have been successfully utilized in a variety of non-murine animals, including sheep, goats, pigs, dogs, cats, monkeys, chimpanzees, hamsters, rabbits, cows and guinea pigs (see, e.g., Kim et al., (1997) Mol Reprod Dev 46:515-526; Houdebine, (1995) Reprod Nutr Dev 35:609-617; Petters (1994) Reprod Fertil Dev 6:643-645; Schnieke et al., (1997) Science 278:2130-2133; and Amoah, (1997) J Animal Science 75:578-585).

The method of introduction of nucleic acid fragments into recombination competent mammalian cells can be by any method which favors co-transformation of multiple nucleic acid molecules. Detailed procedures for producing transgenic animals are readily available to one skilled in the art, including the disclosures in U.S. Patent No. 5,489,743 and U.S. Patent No. 5,602,307.

#### L. Diagnostic Methods

As the genes and proteins of the invention are differentially expressed in cancerous liver tissue (HCC) and in other carcinomas, compared to non-cancerous tissues, the genes and proteins of the invention may be used to diagnose or monitor liver cancer or other malignant neoplasms, to track disease progression, or to differentiate HCC tissue from cirrhotic tissue samples. One means of diagnosing liver cancer using the nucleic acid molecules or proteins of the invention involves obtaining tissue from living subjects, including liver tissue.

The use of molecular biological tools has become routine in forensic technology. For example, nucleic acid probes comprising all or at least part of the sequence of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7 or SEQ ID NO: 9 may be used

10

15

20

25

to determine the expression of a nucleic acid molecule in forensic/pathology specimens. Further, nucleic acid assays may be carried out by any means of conducting a transcriptional profiling analysis. In addition to nucleic acid analysis, forensic methods of the invention may target the proteins of the invention, particularly a protein comprising SEQ ID NO: 2, 4, 6, 8 or 10 to determine up- or down-regulation of the genes (Shiverick et al., (1975) Biochim Biophys Acta 393:124-133).

Methods of the invention may involve treatment of tissues with collagenases or other proteases to make the tissue amenable to cell lysis (Semenov et al., (1987) Biull Eksp Biol Med 104:113-116). Further, it is possible to obtain biopsy samples from different regions of the liver for analysis.

Assays to detect nucleic acid or protein molecules of the invention may be in any available format. Typical assays for nucleic acid molecules include hybridization or PCR based formats. Typical assays for the detection of proteins, polypeptides or peptides of the invention include the use of antibody probes in any available format such as *in situ* binding assays, etc. (see Harlow & Lane, Antibodies - A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1988. In preferred embodiments, assays are carried-out with appropriate controls.

The above methods may also be used in other diagnostic protocols, including protocols and methods to detect disease states in other tissues or organs, for example in tissues in which expression of a nucleic acid molecule of the invention is detected.

Without further description, it is believed that one of ordinary skill in the art can, using the preceding description and the following illustrative examples, make and utilize the compounds of the present invention and practice the claimed methods. The following working examples therefore, specifically point out preferred embodiments of the present invention, and are not to be construed as limiting in any way the remainder of the disclosure.

#### **EXAMPLES**

#### Example 1a

5

10

15

20

25

30

# Identification of Differentially Expressed LBFL302 mRNA in Hepatocellular Carcinoma

The patient tissue samples were derived from 10 Korean patients and classified into two groups of 5 patients each. One group of consisted of patients who had been diagnosed with chronic viral hepatitis B (CH) and who later developed hepatic carcinomas (HCC). The patients in this group, three men and two women, ranged in age from 34-63. The second group of patients had been diagnosed with cirrhosis of the liver (LC). These people also later developed hepatic carcinomas (HCC). In this group of 4 men and one woman, the patients ranged in age from 40-62. For each patient, tissue was obtained from two areas of the liver to produce a set of biopsy samples. In the first patient group (cancer/hepatitis), samples were removed from liver tumors and from the non-cancerous surrounding area composed of inflamed tissue (inflammation due to hepatitis). In the second group (cancer/cirrhosis), liver tissue was removed from tumors and from the non-cancerous surrounding area composed of fibrotic tissue (areas of fibrosis due to cirrhosis). Histological analysis of each of the tissue samples was performed and samples were segregated into either non-cancerous or cancerous categories.

With minor modifications, the sample preparation protocol followed the Affymetrix GeneChip Expression Analysis Manual. Frozen tissue was first ground to powder using the Spex Certiprep 6800 Freezer Mill. Total RNA was then extracted using Trizol (Life Technologies). The total RNA yield for each sample (average tissue weight of 300 mg) was 200-500 μg. Next, mRNA was isolated using the Oligotex mRNA Midi kit (Qiagen). Since the mRNA was eluted in a final volume of 400 μl, an ethanol precipitation step was required to bring the concentration to 1 μg/μl. Using 1-5 μg of mRNA, double stranded cDNA was created using the SuperScript Choice system (Gibco-BRL). First strand cDNA synthesis was primed with a T7-(dT<sub>24</sub>) oligonucleotide. The cDNA was then phenol-chloroform extracted and ethanol precipitated to a final concentration of 1 μg/μl.

From 2 µg of cDNA, cRNA was synthesized according to standard procedures.

10

15

20

25

30

To biotin label the cRNA, nucleotides Bio-11-CTP and Bio-16-UTP (Enzo Diagnostics) were added to the reaction. After a 37 °C incubation for six hours, the labeled cRNA was cleaned up according to the Rneasy Mini kit protocol (Qiagen). The cRNA was then fragmented (5× fragmentation buffer: 200 mM Tris-Acetate (pH 8.1), 500 mM KOAc, 150 mM MgOAc) for thirty-five minutes at 94 °C.

55 μg of fragmented cRNA was hybridized on the Affymetrix Human Genome U95 and U133 set of arrays for twenty-four hours at 60 rpm in a 45 °C hybridization oven. The chips were washed and stained with Streptavidin Phycoerythrin (SAPE) (Molecular Probes) in Affymetrix fluidics stations. To amplify staining, SAPE solution was added twice with an anti-streptavidin biotinylated antibody (Vector Laboratories) staining step in between. Hybridization to the probe arrays was detected by fluorometric scanning (Hewlett Packard Gene Array Scanner). Following hybridization and scanning, the microarray images were analyzed for quality control, looking for major chip defects or abnormalities in hybridization signal. After all chips passed QC, the data was analyzed using Affymetrix Microarray Suite (v4.0), and LIMS (v1.5) for U95 or Affymetrix Microarray Suite (v5.0), and LIMS (v3.0) for U133.

Differential expression of genes between the cancerous and non-cancerous liver samples was determined by using Affymetrix human GeneChip sets, U95 and U133, with the following statistical methods. (1) For each gene, Affymetrix GeneChip average difference values for U95 were determined by Affymetrix Microarray Suite (v4.0), which also made "Absent" (=not detected), "Present" (=detected) or "Marginal" (=not clearly Absent or Present) calls for each GeneChip element. Signal values for U133 were determined by Affymetrix Microarray Suite (v5.0), which also made Absent, Present or Marginal calls. (2) Using the criteria of at least 10% present call in both cancerous and non-cancerous liver samples and at least 40% present call in either cancerous or non-cancerous liver sample groups, a gene set was selected for further analysis. (3) Based on the average difference values of U95 data, the gene set was split into two groups, a high expression group and low expression group. The high expression group contained genes with average difference values greater than or equal to 5 in both cancerous and non-cancerous samples. The remainder of the genes were included in the low

10

15

20

25

30

expression group. The average difference values were transformed to a logarithmic scale for the high expression group, but were not changed for the low expression group. For U133 data, all signal values were transformed to a logarithmic scale regardless of expression level. (4) The Analysis of Variance (ANOVA) method was used for data analysis (Steel et al., Principles and Procedures of Statistics: A Biometrical Approach, Third Ed., McGraw-Hill, 1997). Prior to the final analysis, a leave-one-out approach is used for outlier detection. One sample at a time was left out of the ANOVA analysis to determine whether or not omitting a specific sample from the analysis had any significant effect on the final result. If so, that particular sample was excluded from the final analysis. After outlier detection, a list of genes that are differentially expressed with a p-value of less than or equal to 0.05 was generated by ANOVA. Data from Affymetrix GeneChip U133 chip set was analyzed with a similar procedure. (5) Two additional criteria were used to reduce the number of genes in the gene list generated from U95. Firstly, geometric mean values were compared between the non-cancerous control group samples and the carcinoma disease group samples to obtain a set of genes showing at least 2.0-fold increases or decreases in expression level. Secondly, the ratio of the fold-change value and the p-value had to be 400 or greater.

Analysis of the chip data showed that the expression of the marker LBFL302 was significantly up-regulated (6.2/6-fold, p = 0.00116 for U95; 7.87-fold, p = 0.000944 for U133) in liver carcinoma samples compared to samples from cirrhotic liver tissue. Up-regulation (2.76-fold, p > 0.05 for U95; 4.64-fold, p = 0.0115 for U133) was also observed in liver carcinoma samples compared to tissue samples from inflamed liver tissue (biopsies from areas of inflammation in chronic hepatitis patients). These data indicate that up-regulation of LBFL302 may be diagnostic for liver cancer in people with cirrhosis and may also be diagnostic for liver cancer in people with chronic hepatitis.

The expression level of LBFL302 (SEQ ID NO: 1 or 3) can be measured by chip sequence fragment nos. 51263\_at and 226936\_at on Affymetrix GeneChips® U95 and U133, respectively. The expression levels of 51263\_at and 226936\_at in various malignant neoplasms, compared to normal control tissues, are shown in Table 1a, where the fold-change and the direction of the change (up- or down-regulation) are also

indicated. A fold-change greater than 1.5 was considered to be significant.

Table 1a

					-14	4-40. A
Affy 10.	Tissue	Disease	Merphology	Change	<u> </u>	E 7
	San China Carlo	The second of th	TRANSITIONAL CELL CARCINOMA, NOS	6:1	3	3.9
51263 at	BLADDER, NOS	MALIGNANI NEOPLASM OF BLADDER, NOS	INEI TRATING DUCT CARCINOMA	1.8	3	3.6
226936 at	BREAST, NOS	MALIGNANI NEOPLASM OF FEMALE BARASI, MOS	MILCINOUS ADENOCARCINOMA	-1.7	Down Page	-3.5
51263 at		200	SOLIAMOLIS CELL CARCINOMA, NOS	3.5	<u>9</u>	4.1
226936 at	CERVIX, NOS	VIV.	ADENOCARCINOMA, NOS	2.2	d O	7.8
226936 at	COLON, NOS	MALIGNANI NEOPLASM OF COLON, NOS	MULLERIAN MIXED TUMOR	2.4	<u>a</u>	2.6
226936_at	226936 at ENDOMETRIUM, NOS	MALIGNANI NEOPLASM OF ENDOMETROM	WI MS' TUMOR	14.3	g N	3.2
226936 at	226936 at KIDNEY, NOS	MALIGNAN NEOFLASM OF NUMET, NOS	ADENOCARCINOMA, NOS	5.6	đ	5.7
226936_at			SCHIAMOHS CELL CARCINOMA, NOS	0.4	a n	6.3
226936 at	TUNG, NOS		ADENOCARCINOMA, NOS	1.7	g S	2.2
226936 at		MALIGNANI NEOPLASM OF LONG, NOS	MILLI ERIAN MIXED TUMOR	23	αŊ	4.6
51263 at	IMYOMETRIUM, NOS	SECONDARY MALIGNANI NEURIASM OF MATCHICING	ADENOCARCINOMA NOS	5.7	3	2.3
226936 at	OVARY, NOS	MALIGNANT NEOPLASM OF OVARY	PAPII I ARY SEROUS ADENOCARCINOMA	4.8	<u>s</u>	3.9
226938_at		MALIGNAN I NEOFLASM OF OVARI	ANENOCARCINOMA NOS	3.2	3	2.2
226936 at	226936 at PANCREAS, NOS	MALIGNANT NEOPLASM OF PANCKEAS, NOS	ADENOCARCINOMA NOS	97-	Down O	-5.6
226936_at	226936 at PROSTATE, NOS	MALIGNANI NEOPLASM OF PROSTALE	ADENOCARCINOMA NOS	2.1	윤	6.5
226936 at	RECTUM, NOS	MALIGNANT NEOPLASM OF RECTUM	COLIMACIS CELL CARCINOMA, NOS	2,8	<u>s</u>	3.1
226936 at	SKIN, NOS		DASAI CELL CARCINOMA NOS	12	3	2.9
226936_at		MALIGNANT NEOPLASM OF SKIN, NOS	MALICHANT I VADHOMA NOS	30	3	3.1
51263_at	SMALL INTESTINE, NOS	5	MALIGORIA CITA CARONIONA NOS	3.5	<u> ۽</u>	3.9
226936_at	226936_at SOFT TISSUES, NOS	SECONDARY MALIGNANT NEOPLASM OF CONNECTIVE AND OTHER SOFT TISSUES, NOS	SQUAMOUS CELL CARCINOMA, NOS	3	<u>.</u>	,
226936 at	SPLEEN, NOS	CHRONIC MYELOID LEUKEMIA WITHOUT MENTION OF REMISSION CHRONIC MYELOID LEUKEMIA	CHRONIC MYELOID LEUKEMIA	9:9_	<u>ප</u>	<u>6</u>
			ADENOCARCINOMA, NOS	2.3	3	4.6
226936 at	226936 at STOMACH, NOS	MALIGNANT NEOPLASM OF STOWACH, NOS	SIGNET RING CELL CARCINOMA	1.9	3	2.9
226936_at						

10

15

20

Table 2 summarizes the differential expression data collected from experiments using Affymetrix GeneChips by tissue type. The chips were scanned and the data analyzed by the GX Scan algorithm, which is described in related applications 60/331,182, 60/388,745 and 60/390,608, all entitled "An Automated Computer-based Algorithm for Organizing and Mining Gene Expression Data Derived from Biological Samples with Complex Clinical Attributes," and all of which are herein incorporated by reference in their entirety.

Table 2

LBFL302 is up-regulated in certain types of the following malignant neoplasms with a fold change of 1.5 and above:

	51263 at From U95 data	226936 at From U133 data
1. Bladder	UP	0.00
2. Breast		
(Infiltrating duct cardnomic	a) UP	UP
3. Cervix	UP	UP
4. Colon	UP	UP
5. Kidney	UP	UP
6. Liver	UP	UP
7. Lung	, UP	UP
8. Myometrium	UP.	
9. Ovary	UP	UP
10.Pancreas	UP	UP
11.Rectum	UP	UP
12.Skin	UP	UP
13.Smail Intestine	UP	
14. Soft tissues	UP	-UP
15.Spieen	UP	UP
16. Stomach	UP	UP

The GeneChip expression results, determined by sample binding to chip sequence fragment no. 51263\_at, were validated by quantitative RT-PCR (Q-RT-PCR) using the Taqman® assay (Perkin-Elmer). PCR primers designed from the sequence information file of the specific Affymetrix fragment (51263\_at) were used in the assay. The target gene in each RNA sample (ten ng of total RNA) was assayed relative to an exogenously spiked reference gene. For this purpose, the tetracycline resistance gene was used as the exogenously added spike. This approach provides the relative expression as measured by cycle threshold (Ct) value of the target mRNA relative to a

constant amount of Tet spike Ct values. The sample panel included liver cirrhosis (LC), chronic hepatitis (CH) and hepatocellular carcinoma (HCC) tissue RNAs that were analyzed on U95 GeneChips. In addition, several new samples that were not analyzed on the GeneChip were used for the expression validations by Q-RT-PCR. The Q-RT-PCR data confirms the up-regulation of LBFL302 observed in HCC, compared to LC or CH biopsy samples.

#### Example 1b

5

10

15

20

25

30

#### Identification of Differentially Expressed LBFL303 mRNA in Hepatocellular Carcinoma

The patient tissue samples were derived from 19 Korean patients and classified into two groups. One group of consisted of nine patients who had been diagnosed with chronic viral hepatitis B (CH) and who later developed hepatic carcinomas (HCC). The patients in this group, five men and four women, ranged in age from 34-65. The second group, of ten patients, had been diagnosed with cirrhosis of the liver (LC). These people also later developed hepatic carcinomas (HCC). In this group of eight men and two woman, the patients ranged in age from 37-62. The same procedures as in the above Example 1a were then carried out for each patient.

Analysis of the data from U95 chips showed that expression of the marker corresponding to SEQ ID NO: 5, 7 or 9 was significantly up-regulated (9.34-fold, p-value = 1.44 X 10<sup>-4</sup>) in liver carcinoma samples compared to samples from cirrhotic liver tissue. Data from U133 chips showed that expression of SEQ ID NO: 5 or 7 was also significantly up-regulated in liver carcinoma samples compared to samples from cirrhotic liver tissue (2.60-fold, p-value = 3.63 X 10<sup>-2</sup>) and compared to samples from inflamed liver tissue (5.69-fold, p = 8.99 X 10<sup>-4</sup> for U95) (biopsies from areas of inflammation in chronic hepatitis patients). These data indicate that up-regulation of SEQ ID NOS: 5, 7 and 9 may be diagnostic for liver cancer in people and, in particular, patients with cirrhosis or chronic hepatitis.

The expression level of LBFL303 clones GE6, MB5 or IE4 (SEQ ID NOS: 5, 7 or 9, respectively) can be measured by chip sequence fragment nos. 46690\_at (U95 chip) and 219175\_at and 224931\_at (U133 chip) on Affymetrix GeneChips®. Differential

10

expression data were collected from experiments using Affymetrix GeneChips® by tissue type and were analyzed by the GX Scan algorithm, which is described in related applications 60/331,182, 60/388,745 and 60/390,608, all entitled "An Automated Computer-based Algorithm for Organizing and Mining Gene Expression Data Derived from Biological Samples with Complex Clinical Attributes," and all of which are herein incorporated by reference in their entirety. The expression levels of 46690\_at, 219175\_at and 224931\_at in various malignant neoplasms, compared to normal control tissues, are shown in Table 1, where the fold-change and the direction of the change (upor down-regulation) are also indicated. A fold-change greater than 1.5 was considered to be significant.

Table 1b

366,10	Tikking	(Discharge Value of the Control of t	Morphology	Fold	DIK TESTAT	T-Stat
<u>.</u>			and the second of the second o	Change		
		MALIGNANT NEOPLASM OF FEMALE BREAST, NOS	INFILTRATING DUCT AND LOBULAR			
46690 at	BREAST, NOS		CARCINOMA	1.7	Down	-5.5
46690 at	CERVIX NOS	MALIGNANT NEOPLASM OF UTERINE CERVIX	SQUAMOUS CELL CARCINOMA, NOS	-1.7	Down	4.5
224931 at	KIDNEY NOS	MALIGNANT NEOPLASM OF KIDNEY, NOS	WILMS' TUMOR	1.6	9	6.3
46590 at		MAI IGNANT NEOPLASM OF OVARY	MUCINOUS CYSTADENOCARCINOMA, NOS	-1.5	Down 1-2.9	2.9
219175 s at	OVARY, NOS	MAI IGNANT NEOPLASM OF OVARY	MULLERIAN MIXED TUMOR	-1.6	Down	-11.3
240175 0 at	210175 e at IDANCREAS NOS	MALIGNANT NEOPLASM OF ISLETS OF LANGERHANS	ISLET CELL CARCINOMA	-2.2	Down	-9.2
2	2000	MALIGNANT NEOPLASM OF CONNECTIVE AND OTHER SOFT	FIBROUS HISTIOCYTOMA, MALIGNANT			
46690 at		TISSUES NOS		-1.6	Down -3.2	32
		MALIGNANT NEOPLASM OF CONNECTIVE AND OTHER SOFT	LIPOSARCOMA, NOS			
46690 at	SOFI HISSUES, NOS	TISSUES. NOS		9	L DOWN	-2.6
		MALIGNANT NEOPLASM OF CONNECTIVE AND OTHER SOFT	LIPOSARCOMA, NOS			·
219175 s at		TISSUES, NOS			Down -2.5	22
ARRAO at		IMALIGNANT NEOPLASM OF TESTIS, NOS	MIXED GERM CELL TUMOR		Down -2.8	2.8
46690 at	TESTIS NOS	MALIGNANT NEOPLASM OF TESTIS, NOS	SEMINOMA, NOS	-2.5	Down -2.3	23
224931 at		MAI IGNANT NEOPLASM OF TESTIS, NOS	MIXED GERM CELL TUMOR	-1.7	Down -2.3	2.3
		MALIGNANT LYMPHOMA, NOS OF UNSPECIFIED, EXTRANODAL	MALIGNANT LYMPHOMA, NOS	i		
224931 at	224931 at THYROID GLAND, NOS	OR SOLID ORGAN SITE		-1.7	Down  -4.4	4.
	١					

The GeneChip expression results, determined by sample binding to chip sequence fragment no. 46690\_at, were validated by quantitative RT-PCR (Q-RT-PCR) using the Taqman® assay (Perkin-Elmer), as in the above Example 1a.

The Q-RT-PCR data confirms the up-regulation of the genes corresponding to LBFL303 clones GE6, MB5 and IE4 observed in HCC, compared to LC or CH biopsy samples.

#### 5 Example 2

10

15

20

.25

30

Cloning of Full Length human cDNAs Corresponding to Differentially Expressed mRNA

Species

The full length cDNA having SEQ ID NO: 1, 3, 5, 7 or 9 was obtained by the oligo-pulling method. Briefly, a gene-specific oligo was designed based on the sequence of SEQ ID NO: 1, 3, 5, 7 or 9. The oligo was labeled with biotin and used to hybridize with 2 µg of single strand plasmid DNA (cDNA recombinants) from a fully differentiated stomach adenocarcinoma library (NCI CGAP Gas 4) following the procedures of Sambrook et al. The hybridized cDNAs were separated by streptavidin-conjugated beads and eluted by heating. The eluted cDNA was converted to double strand plasmid DNA and used to transform E. coli cells (DH10B) and the longest cDNA was screened. After positive selection was confirmed by PCR using gene-specific primers, the cDNA clone was subjected to DNA sequencing.

The nucleotide sequence of the full-length human cDNAs corresponding to the differentially regulated mRNA detected above is set forth in SEQ ID NOS: 1, 3, 5, 7 and 9. The cDNA of SEQ ID NO: 1 comprises 578 base pairs (531 base pairs and a polyA tail), and the cDNA of SEQ ID NO: 3 comprises 531 base pairs (515 base pairs and a polyA tail). The cDNA of SEQ ID NO: 5 comprises 2067 base pairs (2040 base pairs and a polyA tail), the cDNA of SEQ ID NO: 7 comprises 2178 base pairs (2162 base pairs and a polyA tail), and the cDNA of SEQ ID NO: 9 comprises 1616 bases pairs (1598 base pairs and a polyA tail).

An open reading frame within the cDNA nucleotide sequence of SEQ ID NO: 1, at nucleotides 155-418 (155-421 including the stop codon), encodes a protein of 88 amino acids. The amino acid sequence corresponding to a predicted protein encoded by SEQ ID NO: 1 is set forth in SEQ ID NO: 2.

An open reading frame within the cDNA nucleotide sequence of SEQ ID NO: 3,

10

15

20

25

at nucleotides 139-402 (139-405 including the stop codon), also encodes a protein of 88 amino acids. The amino acid sequence corresponding to a predicted protein encoded by SEQ ID NO: 3 is set forth in SEQ ID NO: 4. The protein sequence of SEQ ID NO: 4 is identical to that of SEQ ID NO: 2, except for the amino acid at position 28 (arginine in SEQ ID NO: 2, but leucine in SEQ ID NO: 4, caused by a G->T point mutation at nucleotide position no. 237 in SEQ ID NO: 1 and 221 in SEQ ID NO: 3), although the nucleic acid sequences encoding these proteins differ upstream of the coding region.

SEQ ID NOS: 2 and 4 are weakly similar to histone-like transcription factor (CBF/NF-Y) and the archaeal histone signature. In addition, these amino acid sequences are weakly similar to the bacterial regulatory protein lysR family helix-turnhelix signature. This signature contains three domains. The amino acid sequences of SEQ ID NO: 2 and SEQ ID NO: 4 are 22% identical to the two domains at the C-terminus.

An open reading frame within the cDNA nucleotide sequence of SEQ ID NO: 5 (GE6), at nucleotides 32-1384 (32-1387 including the stop codon), encodes a protein of 451 amino acids. The amino acid sequence corresponding to the protein encoded by SEQ ID NO: 5 is set forth in SEQ ID NO: 6.

An open reading frame within the cDNA nucleotide sequence of SEQ ID NO: 7 (MB5), at nucleotides 41-1501 (41-1504 including the stop codon), encodes a protein of 487 amino acids. The amino acid sequence corresponding to the protein encoded by SEQ ID NO: 7 is set forth in SEQ ID NO: 8. The protein sequence of SEQ ID NO: 8 is identical to that of SEQ ID NO: 6, except for an insertion of 36 amino acids toward the amino terminus (see multiple sequence alignment below).

An open reading frame within the cDNA nucleotide sequence of SEQ ID NO: 9 (IE4), at nucleotides 31-1551 (31-1554 including the stop codon), encodes a protein of 507 amino acids. The amino acid sequence corresponding to the protein encoded by SEQ ID NO: 9 is set forth in SEQ ID NO: 10. The protein corresponding to SEQ ID NO: 10 is identical to that of SEQ ID NO: 8 for the first 456 amino acids, although the remainders of the proteins have no homology (see multiple sequence alignment below).

#### Multiple amino acid sequence alignment for GE6, MB5 and IE4

LBFL303_GR6.pep LBFL303_MR5.pep LBFL303_IR4.pep Consensus	1 50 MDGTETRQRR LDSCGKPGEL GLPHPLSTGG LPVASEDGAL RAPESQSVTP
LBFL303_GE6.pep LBFL303_MB5.pep LBFL303_IE4.pep Consensus	KPLETEPSRE TAWSIGLQVT VPFMFAGLGL SWAGMLLDYF Q KPLETEPSRE TAWSIGLQVT VPFMFAGLGL SWAGMLLDYF QHWPVFVEVK KPLETEPSRE TAWSIGLQVT VPFMFAGLGL SWAGMLLDYF QHWPVFVEVK KPLETEPSRE TAWSIGLQVT VPFMFAGLGL SWAGMLLDYF QHWPVFVEVK
LBFL303_GE6.pep LBFL303_MB5.pep LBFL303_IE4.pep Consensus	150ANT GQIDDPQEQH RVISSNLALI DLLTLVPPLV GLKGNLEMTL ASRLSTAANT GQIDDPQEQH RVISSNLALI DLLTLVPPLV GLKGNLEMTL ASRLSTAANT GQIDDPQEQH RVISSNLALI DLLTLVPPLV GLKGNLEMTL ASRLSTAANT GQIDDPQEQH RVISSNLALI
LBFL303_GE6.pep LBFL303_MB5.pep LBFL303_IE4.pep Consensus	200 QVQATVVGLL AAVAALLIGV VSREEVDVAK VELLCASSVL TAFLAAFALG
LBFL303_GE6.pep LBFL303_MB5.pep LBFL303_IE4.pep Consensus	250 VLMVCIVIGA RKLGVNPDNI ATPIAASLGD LITLSILALV SSFFYRHKDS
LBPL303_GE6.pep LBPL303_MB5.pep LBPL303_IE4.pep Consensus	251 300 RYLTPLVCLS FAALTEVWVL IAKQSPPIVK ILKFGWFPII LAMVISSFGG
LBFL303_GE6.pep LBFL303_MB5.pep LBFL303_IE4.pep Consensus	350 LILSKTVSKQ QYKGMAIFTP VICGVGGNLV AIQTSRISTY LHMWSAPGVL
LBFL303_GE6.pep LBFL303_MB5.pep LBFL303_IR4.pep Consensus	400 PLOMKKPWPN PCSTFCTSEI NSMSARVLLL LVVPGHLIFF YIIYLVEGQS PLOMKKFWPN PCSTFCTSEI NSMSARVLLL LVVPGHLIFF YIIYLVEGQS PLOMKKFWPN PCSTFCTSEI NSMSARVLLL LVVPGHLIFF YIIYLVEGQS PLOMKKFWPN PCSTFCTSEI NSMSARVLLL LVVPGHLIFF YIIYLVEGQS
LBFL303_GE6.pep LBFL303_MB5.pep LBFL303_IE4.pep Consensus	401 VINSQTFVVL YLLAGLIQVT ILLYLAEVMV RLTWHQALDP DNHCIPYLTG
LBFL303_GR6.pep LBFL303_MB5.pep LBFL303_IE4.pep Consensus	451 LGDLLGTGLL ALCPFTDWLL KSKARLGGIS BLASGPP* LGDLLGTGLL ALCFFTDWLL KSKARLGGIS BLASGPP* LGDLLGSSSV GHTAAVPRRC TASPGWGLIQ PFICTOHLIV SLLSFYFPPC LGDLLGTGLL ALCFFTDWLL KSKARLGGIS BLASGPP
LBFL303_GE6.pep LBFL303_NB5.pep LBFL303_IE4.pep Consensus	LLAKTSI+

The LBFL303 clones also exhibit partial homology to a mouse homologue (GenBank Accession No. XM\_132686). SEQ ID NO: 5 shows 44% identity over the entire contiguous sequence and 68% identity within the open reading frame, while SEQ

10

15

20

25

30

ID NO: 7 shows 46% identity over the entire contiguous sequence and 69% identity within the open reading frame. SEQ ID NO: 9 shows 61% identity over the entire contiguous sequence and 64% identity within the open reading frame to the mouse nucleic acid sequence.

SEQ ID NOS: 6, 8 and 10 contain a divalent cation transporter signature that has two domains. SEQ ID NOS: 6 and 8 contain both domains (at amino acid positions 85-205 and 282-428 in SEQ ID NO: 6, and at amino acid positions 105-241 and 318-464 in SEQ ID NO: 8), although SEQ ID NO: 10 contains only the N-terminal domain (from amino acid residues 105-241). In addition, SEQ ID NOS: 6, 8 and 10 contain a peroxidase-active site at amino acid positions 390-401 in SEQ ID NO: 6 and at amino acid positions 426-437 in SEQ ID NOS: 8 and 10.

Figures 1, 2, 3, 4 and 5 show the results of a hydrophobicity analysis of the amino acid sequence of SEQ ID NOS: 2, 4, 6, 8 and 10. Hydrophilic regions may be used to produce antigenic peptides, as described above.

Analysis by Northern blot was performed to determine the size of the mRNA transcripts that correspond to SEQ ID NOS: 1, 3, 5, 7 and 9. Northern blots containing total RNAs from various human tissues were used (ClonTech), and each of clone BC7 (SEQ ID NO: 3), clone GE6 (SEQ ID NO: 5), clone MB5 (SEQ ID NO: 7), clone IE4 (SEQ ID NO: 9) was radioactively labeled by the random primer method and used to probe the blots. The blots were hybridized in Church and Gilbert buffer at 65 °C and washed with 0.1X SSC containing 0.1% SDS at room temperature. The Northern blots show a single transcript for each LBFL302 and 303 gene, which is approximately 0.65 kb and 2.45 kb in size. This corresponds to the size of the insert in clones BC7 and BC4 (SEQ ID NOS: 3 and 1), 0.531 kb and 0.578 kb, respectively, and 2.2 kb for GE6, 2.3 kb for MB5, and 1.8 kb for IE4, SEQ ID NOS: 5, 7 and 9, respectively.

To examine the expression of SEQ ID NO: 1, 3, 5, 7 or 9 in various normal tissues, an electronic Northern blot (e-Northern) was prepared as follows. Using the chips and the procedures in Example 1, mRNA from a panel of 46 normal tissues, as listed in Table 3, was hybridized to Affymetrix U95 human GeneChips. The results of these experiments are shown in Table 3. For each tissue type, the number of samples that are called

present or absent are indicated, together with the total number of samples in that sample set. In addition, the median value and the 25<sup>th</sup> and 75<sup>th</sup> percentiles in each tissue type are listed. Interestingly, although this gene is up-regulated in liver cancer, expression of LBFL302 and 303 could not be detected in most normal liver samples. This observation indicates that LBFL302 and 303 may be used as a diagnostic agent or marker to detect liver cancer or to differentiate hepatocellular carcinoma from cirrhotic liver tissue, as discussed below. Expression levels of LBFL302 appeared to be highest in the thymus, followed by organs of the reproductive system (testis, endometrium, myometrium, uterus, cervix and breast) and of the digestive system (esophagus, rectum, colon and appendix). Expression levels of LBFL303 appeared to be highest in the brain (cerebellum, frontal cortex, temporal cortex and hippocampus) and in organs of the reproductive system (testis, endometrium, myometrium, uterus, cervix and breast), but expression could be detected at lower levels in most other tissues. Expression in the liver was the lowest.

10

5

Table 3a- e-Northern Table for 51263\_at: LBFL302 Gene Expression in Normal Tissues

Global	Tissue	Present	Absent ·	Lower 25%	Median	Upper 75%
Present	1	į	· ·	1		
Freq.	<u> </u>					
0.7176	<u> </u>				·	
	Adipose	26 of 32	6 of 32	32.78	40.25	56.11
	Adrenal Gland	5 of 12	7 of 12	10.18	28.24	42.42
	Appendix	3 of 3	0 of 3	65.56	75.82	78.15
	Artery	3 of 3	0 of 3	39.18	43.53	48.12
	Bladder	.5 of 5	0 of 5	49.95	81.92	89.71
	Bone	3 of 3	0 of 3	73.57	112.71	113.59
	Breast	76 of 80	4 of 80	56.53	79.61	107.84
<del>,</del>	Cerebellum	1 of 5	4 of 5	10 .58	16.98	17.79
	Cervix	96 of 101	5 of 101	75.45	102.71	139.70
	Colon	147 of 151	4 of 151	72.44	107.19	149.44
	Cortex Frontal Lobe	2 of 7	5 of 7	11.01	17.48	21.95
	Cortex Temporal	0 of 3	3 of 3	-1.09	10.90	11.83
	Lobe					<u> </u>
	Duodenum	58 of 61	3 of 61	57.59	72.35	94.63
	Endometrium	21 of 21	0 of 21	123.36	151.44	185.97
	Esophagus	26 of 27	1 of 27	98.53	127.38	167.07
	Fallopian Tube	46 of 51	5 of 51	37.81	52.34	94.69
	GaliBladder	2 of 8	6 of 8	13.23	32.72	39.09
	Heart	l of 3	2 of 3	8.24	13.13	15.87
	Hippocampus	2 of 5	3 of 5	17.28	19.47	31.10
	Kidney	24 of 86	62 of 86	8.63	14.36	20.22
	Larynx	4 of 4	0 of 4	40.57	97.82	163.75
	Left Atrium	30 of 141	111 of 141		14.57	19.93
	Left Ventricle	0 of 15	15 of 15	4.77	8.67	12.80
<del></del>	Liver	10 of 34	24 of 34	2.08	10.42	19.94
	Lung	62 of 93	31 of 93	23.04	35.46	55.34
	Lymph Node	10 of 11	1-of 11	58.47	85.56	98.77
	Muscles	9 of 39	30 of 39	9.03	16.64	. 27.86
	Myometrium	105 of 106	1 of 106	123.43	180.91	216.99
·	Omentum	14 of 15	1 of 15	46.67	67.85	151.98
	Ovary	49 of 74	25 of 74	24.22	34.54	55.81
	Pancreas	2 of 34	32 of 34	-2.88	6.00	14.34
	Placenta	3 of 5	2 of 5	28.65	40.79	53.18
	Prostate	30 of 32	2 of 32	44.39	59.04	.73.73
	Rectum	42 of 43	1 of 43	97.69	133.62	173.02
	Right Atrium	34 of 169	135 of 169		14.77	19.39
	Right Ventricle	21 of 160	139 of 160		10.56	18.22
L	Skin	50 of 59	9 of 59	44.51	64.27	93.29
	Small Intestine	60 of 68	8 of 68	41.12	70.61	103.24
	Soft Tissues	4 of 6	2 of 6	26.64	61.11	115.92
	Spieen	22 of 31	9 of 31	26.94	39.23	52.19
	Stomach	34 of 47	13 of 47		41.88	80.51
	Testis	5 of 5	0 of 5	137.44	236.97	359.92
	Thymus	71 of 71	0 of 71	261.31	322.22	
	Thyroid Gland	6 of 18	12 of 18	10.27	15.12	31.99
	Uterus	58 of 58	0 of 58	88.62	140.06	190.83
	WBC	12 of 40	28 of 40	10.23	15.37	24.59

Table 3b- e-Northern Table for 46690\_at: LBFL303 Gene Expression in Normal Tissues

Global				·		
Present	Tissue	Present	Absent	1 auga 250/	34	
Freq.		· rosone	Vnaciif	Lower 25%	Median	Upper 75%
0.6295						<del></del>
	Adipose	21 of 32	11 of 32	176.42	227.48	
	Adrenal Gland	10 of 12	2 of 12	195.93	286.51	291.00
	Appendix	2 of 3	1 of 3	59.10	61.20	364.93
	Artery	3 of 3	0 of 3	257.57	268.59	71.28
	Bladder	4 of 5	1 of 5	222.21	236.51	334.19
	Bone	2 of 3	1 of 3	241.64	263.23	310.37
	Breast	74 of 80	6 of 80	208.22	279.85	269,74
	Cerebellum	5 of 5	0 of 5	333.02	348.16	332.28 421.84
	Cervix	88 of 101	13 of 101	212.62	292.96	
	Colon	70 of 151	81 of 151	63.11	98.21	344.12
	Cortex Frontal Lobe	7 of 7	0 of 7	293.14	313.78	142.28
	Cortex Temporal Lobe	3 of 3	0 of 3	255.54	260.80	323.67 339.41
	Duodenum	31 of 61	30 of 61	86.34	113.58	
	Endometrium	18 of 21	3 of 21	243.49	259.81	143.09 299.21
	Esophagus	17 of 27	10 of 27	120.68	166.05	207.75
	Fallopian Tube	45 of 51	6 of 51	236.80	296.76	358.36
	GallBladder	8 of 8	0 of 8	250.23	277.49	<u> </u>
	Heart	2 of 3	1 of 3	128.08	129.40	328.73
	Hippocampus	5 of 5	0 of 5	216.82	302.36	139.86 337.49
•	Kidney	53 of 86	33 of 86	155.75	202.37	235.42
	Larynx	3 of 4	1 of 4	189.07	222.38	268.94
	Left Atrium	75 of 141	66 of 141	130.83	173:47	220.14
	Left Ventricle	8 of 15	7 of 15	117.45	175.84	208.91
·	Liver	5 of 34	29 of 34	8.65	33.49	55,39
	Lung	52 of 93	41 of 93	115.09	167.11	232.69
· · ·	Lymph Node	8 of 11	3 of 11	119.03	167.95	223.28
	Muscles	32 of 39	7 of 39	147.91	200.14	226.77
	Myometrium	98 of 106	8 of 106	248.04	328.05	407.02
	Omentum	8 of 15	7 of 15	88.63	179.55	196.14
	Ovary	71 of 74	3 of 74	306.34	384.95	435.31
	Pancreas	10 of 34	24 of 34	78.67	128.72	162.20
	Placenta	1 of 5	4 of 5	79.06	108.81	112.22
	Prostate	23 of 32	9 of 32	181.76	193.51	230.41
	Rectum	28 of 43	15 of 43	113.19	126.40	
	Right Atrium	82 of 169	87 of 169	134.58	166.54	
	Right Ventricle	100 of 160	60 of 160	122.83	165.44	
	Skin	49 of 59	10 of 59	185.75		297.31
	Small Intestine	33 of 68	35 of 68	69.62	109.95	
	Soft Tissues	5 of 6	1 of 6	186.44	214.50	
	Spleen	22 of 31	9 of 31	149.56	184.29	
	Stomach	20 of 47	27 of 47	77.44	114.84	+
	Testis	5 of 5	0 of 5	396.28		

Global Present Freq.	Tissue	Present	Absent	Lower 25%	Median	Upper 75%
	Thymus	70 of 71	1 of 71	296,44	348.71	400.48
	Thyroid Gland	13 of 18	5 of 18	228.07	310.09	350.55
	Uterus	53 of 58	5 of 58	249.90	318.35	361.58
	WBC	15 of 40	25 of 40	68.93	94.14	122.44

#### INDUSTRIAL APPLICABILITY

#### 5 Example 3

10

15

20

#### Detection of LBFL302 and 303 mRNA for Liver Cancer Screening

The expression level of mRNA corresponding to SEQ ID NO: 1, 3, 5, 7 or 9 is determined in liver tissue biopsy samples, as described in Example 1, *i.e.*, by screening mRNA samples on a GeneChip, or as described in Example 2, *i.e.*, by screening mRNA samples on a Northern blot. Alternatively, samples from non-liver hyperplastic tissues in malignant or non-malignant states may also be analyzed. Liver tissue samples from patients with liver cancer and from normal subjects may be used as positive and negative controls. Using any means of assaying gene expression, a level of expression higher than that of the normal control is indicative of liver cancer or a likelihood of developing liver cancer.

Although the present invention has been described in detail with reference to examples above, it is understood that various modifications can be made without departing from the spirit of the invention. Accordingly, the invention is limited only by the following claims. All cited patents, patent applications and publications referred to in this application are herein incorporated by reference in their entirety.

T/KR2003/001655

#### What Is Claimed:

5

10

15

- 1. An isolated nucleic acid molecule selected from the group consisting of: (a) an isolated nucleic acid molecule comprising SEQ ID NO: 1, 3, 5 or 7, (b) an isolated nucleic acid molecule encoding SEQ ID NO: 6 or 8, (c) an isolated nucleic acid molecule that encodes a protein that is expressed in liver cancer and that exhibits at least about 95% nucleotide sequence identity over the entire contiguous sequence of SEQ ID NO: 5, (d) an isolated nucleic acid molecule that encodes a protein that is expressed in liver cancer and that exhibits at least about 75% nucleotide sequence identity over the entire contiguous sequence of SEQ ID NO: 7, and (e) an isolated nucleic acid molecule comprising the complement of a nucleic acid molecule of (a), (b), (c) or (d).
- 2. The isolated nucleic acid molecule of claim 1, wherein the nucleic acid molecule consists of nucleotides 155-418 of SEO ID NO: 1.
- 3. The isolated nucleic acid molecule of claim 1, wherein the nucleic acid molecule consists of nucleotides 139-402 of SEQ ID NO: 3.
- 4. The isolated nucleic molecule of claim 1, wherein the nucleic acid molecule comprises nucleotides 32-1384 of SEQ ID NO: 5.
  - 5. The isolated nucleic molecule of claim 1, wherein the nucleic acid molecule comprises nucleotides 32-1387 of SEQ ID NO: 5.
- 25 6. The isolated nucleic acid molecule of claim 1, wherein the nucleic acid molecule consists of nucleotides 32-1384 of SEQ ID NO: 5.
  - 7. The isolated nucleic molecule of claim 1, wherein the nucleic acid molecule comprises nucleotides 41-1501 of SEQ ID NO: 7.

- 8. The isolated nucleic molecule of claim 1, wherein the nucleic acid molecule comprises nucleotides 41-1504 of SEQ ID NO: 7.
- 9. The isolated nucleic acid molecule of claim 1, wherein the nucleic acid molecule consists of nucleotides 41-1501 of SEQ ID NO: 7.
  - 10. The isolated nucleic acid molecule of any one of claims 1-9, wherein said nucleic acid molecule is operably linked to one or more expression control elements.
- 10 11. A vector comprising an isolated nucleic acid molecule of any one of claims 1-9.
  - 12. A host cell transformed to contain the nucleic acid molecule of any one of claims 1-9.
- 15 13. A host cell comprising a vector of claim 11.
  - 14. A host cell of claim 13, wherein said host cell is selected from the group consisting of prokaryotic host cells and eukaryotic host cells.
- 20 15. A method for producing a polypeptide comprising culturing a host cell transformed with the nucleic acid molecule of any one of claims 1-9 under conditions in which the protein encoded by said nucleic acid molecule is expressed.
- 16. The method of claim 15, wherein said host cell is selected from the group consisting of prokaryotic host cells and eukaryotic host cells.
  - 17. An isolated polypeptide produced by the method of claim 15.
- 18. An isolated polypeptide or protein selected from the group consisting of an isolated polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 4, 6 or 8, an isolated

polypeptide comprising a fragment of at least 10 amino acids of SEQ ID NO: 2 or 4, an isolated polypeptide comprising conservative amino acid substitutions of SEQ ID NO: 2 or 4, an isolated polypeptide comprising naturally occurring amino acid sequence variants of SEQ ID NO: 2 or 4, an isolated polypeptide exhibiting at least about 75% amino acid sequence identity with SEQ ID NO: 2 or 4, and a protein having at least about 95% amino acid sequence identity with SEQ ID NO: 6 or 8.

- 19. An isolated antibody or antigen-binding antibody fragment that binds to a polypeptide of claim 18.
- 20. An antibody of claim 19, wherein said antibody is a monoclonal or a polyclonal antibody.
- 21. A method of identifying an agent which modulates the expression of a nucleic acid encoding a protein of claim 18, comprising:

exposing cells which express the nucleic acid to the agent; and

determining whether the agent modulates expression of said nucleic acid, thereby identifying an agent which modulates the expression of a nucleic acid encoding the protein.

20

25

5

10

15

22. A method of identifying an agent which modulates the level of or at least one activity of a protein of claim 18 or a protein comprising SEQ ID NO: 10, comprising:

exposing cells which express the protein to the agent;

determining whether the agent modulates the level of or at least one activity of said protein, thereby identifying an agent which modulates the level of or at least one activity of the protein.

- 23. The method of claim 22, wherein the agent modulates one activity of the protein.
- 30 24. A method of identifying binding partners for a protein of claim 18 or a protein

comprising SEQ ID NO: 10, comprising:

exposing said protein to a potential binding partner; and

determining if the potential binding partner binds to said protein, thereby identifying binding partners for the protein.

5

25. A method of modulating the expression of a nucleic acid encoding a protein of claim 18 or a protein comprising SEQ ID NO: 10, comprising:

administering an effective amount of an agent which modulates the expression of a nucleic acid encoding the protein.

10

26. A method of modulating at least one activity of a protein of claim 18 or a protein comprising SEQ ID NO: 10, comprising:

administering an effective amount of an agent which modulates at least one activity of the protein.

15

20

- 27. A non-human transgenic animal modified to contain a nucleic acid molecule of any of claims 1-9 or SEQ ID NO: 10.
- 28. The transgenic animal of claim 27, wherein the nucleic acid molecule contains a mutation that prevents expression of the encoded protein.
  - 29. A method of diagnosing a disease state in a subject, comprising:

determining the level of expression of a nucleic acid molecule or protein of any one of claims 1-9 or 18, a nucleic acid comprising SEQ ID NO: 9 or a protein molecule comprising SEQ ID NO: 10.

- 30. The method of claim 29, wherein the disease state is liver cancer.
- 31. The method of claim 30, wherein the disease state is hepatocellular carcinoma.

25

10

15

- 32. The method of claim 29, wherein the disease state is a malignant neoplasm.
- 33. The method of claim 32, wherein the malignant neoplasm occurs in the bladder, breast, cervix, colon, kidney, lung, myometrium, ovary, pancreas, prostate, rectum skin, small intestine, soft tissue, spleen, stomach, testis or thyroid gland.
- 34. A composition comprising a diluent and a polypeptide or protein selected from the group consisting of an isolated polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 4, 6, 8 or 10, an isolated polypeptide comprising a fragment of at least 10 amino acids of SEQ ID NO: 2, 4, 6, 8 or 10, an isolated polypeptide comprising conservative amino acid substitutions of SEQ ID NO: 2 or 4, an isolated polypeptide comprising naturally occurring amino acid sequence variants of SEQ ID NO: 2 or 4, an isolated polypeptide exhibiting at least about 75% amino acid sequence identity with SEQ ID NO: 2 or 4, and a polypeptide exhibiting at least about 95% amino acid sequence identity with SEQ ID NO: 6, 8 or 10.

Fig. 1

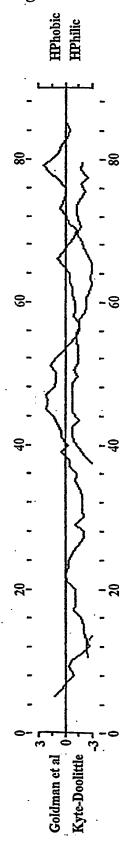
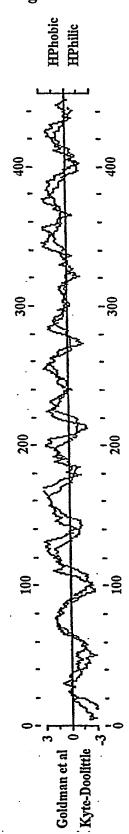


Fig. 2



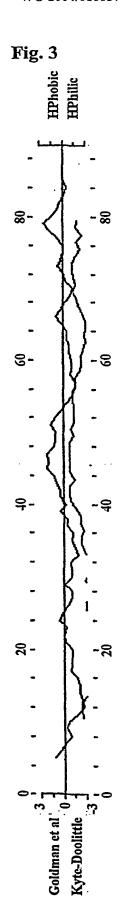


Fig. 4

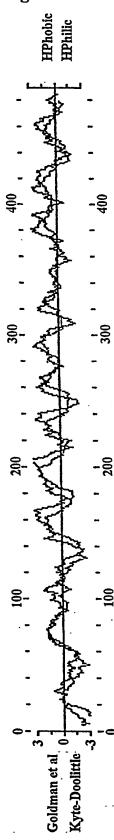
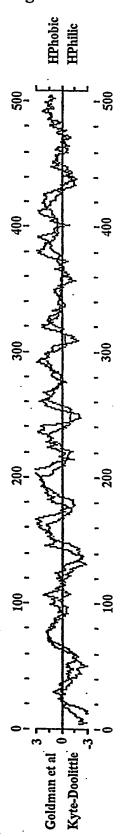
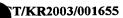
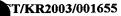


Fig. 5





<110>	LG Life Sciences, Ltd.
<120>	Gene families associated with liver cancer
<130>	PC03016-LG
<150>	US 60/402,905
<151>	2002-08-14
<150>	US 60/403,651
<151>	2002-08-16
<160>	10
<170>	KopatentIn 1.71
<210>	1
<211>	578
<212>	DNA
<213>	Homo sapiens
<220>	
<221>	CDS
<222>	(155)(418)
<223>	Gene LBFL302, Clone BC4
<400>	1
cggacgcg	tg ggttcgaacg ttcggactga ggtttttctg cctgaagaag cgtcatacgg 60
accggatt	gt tttcgctggc ccagtgtccc cggagcttgt gtgcgataca gagagcacct 120
cggaagct	ga ggcagctggt acttgacaga gagg atg gcg ctg tcg acc 169
	Met Ala Leu Ser Thr
	1 5
ata gtc	tcc cag agg aag cag ata aag cgg aag gct ccc cgt ggc ttt 217
	Ser Gln Arg Lys Gln Ile Lys Arg Lys Ala Pro Arg Gly Phe
	10 15 20



cta	aag	cga	gtc	ttc	aag	cga	aag	aag	cct	caa	ctt	cgt	ctg	gag	aaa	265	
Leu	Lys	Arg	Val	Phe	Lys	Arg	Lys	Lys	Pro	Gln	Leu	Arg	Leu	Glu	Lys	•	
			25					30					35				
agt	ggt	gac	tta	ttg	gtc	cat	ctg	aac	tgt	tta	ctg	ttt	gtt	cat	cga	313	
Ser	Gly	Asp	Leu	Leu	Val	His	Leu	Asn	Суз	Leu	Leu	Phe	Val	His	Arg		
		40					45					50					
tta	gca	gaa	gag	tcc	agg	aca	aac	gct	tgt	gcg	agt	aaa	tgt	aga	gtc	361	
Leu	Ala	Glu	Glu	Ser	Arg	Thr	Asn	Ala	Суз	Ala	Ser	ГÀЗ	Cys	Arg	Val	•	
	55					60					65						
att	aac	aag	gag	cat	gta	ctg	gcc	gca	gca	aag	gta	att	cta	aag	aag	409	
Ile	Asn	Lys	Glu	His	Val	Leu	Ala	Ala	Ala	Lys	Val	Ile	Leu	Lys	Lys		
70					75					80					85		
agc	aga	ggt		•	ta g	aagt	caaaq	g aa	cata	ttct	tga	aagti	tat	gatg	cattct	460	1
Ser	Arg	Gly															
ttt	gggt	ggt	aaca	gate	at a	aaga	catt	t tt	taca	catc	agt	taata	atg	ggat	tattaa	520	,
																_	
ata	ttgg	cta	taaa	aaaa	aa a	aaaa	aaaa	a aa	aaaa	aaaa	aaa	aaaa	aaa	aaaa	aaaa	578	
		_															
<21		2															
<21		88															
<21		PRT 															
<21	3>	ното	sap	iens													
-40	٥.	^															
<40		2	<b>a</b>	_,				<b>.</b>	•		<b>01</b> =	~1-	<b>.</b>	<b>3</b>	~		
		Leu	Ser			Val	ser	GIN		_	GIN	TTE	гÀг	-	- ·.		
1				. 5					10					15			
<b>71</b> -	D	3		- DL-	•	•		**- 3	75 a	T	70		T	D	C1		
ATS	Pro	Arg	_		Leu	груs	Arg			: гуда	Arg	гÃг	_		Gln		
			20	1				25					30				
T	<b>n</b>			<b></b>			. n	T	T		n 4 -	T =	7	· C	Ton		
nen	Arg			г гла	ser	GTĀ	•		теп	val	nis			cys	Leu		
		35	,				40					45	,				
7	DL -	17~1	<b>.</b>	70				C1			መኤ~		- רת	· ~··-	. NI-		
пеn	rne	. Ag1	. пт2	MIG	, res	r WTS	GIU	GIU	ser	. Arg	TIII	. ASD	. WTS	. Cys	Ala		

### Sequence Listing

50	55	60	
Ser Lys Cys Arg 65	Val Ile Asn Lys Glu 70	His Val Leu Ala A	la Ala Lys 80
Val Ile Leu Lys	Lys Ser Arg Gly 85		
<210> 3 <211> 531 <212> DNA <213> Homo s	sapiens		•
<220> <221> CDS <222> (139)	(402)		
<223> Gene 1	LBFL302, Clone BC7		
<400> 3 cccacgcgtc cgg	aggtttt tctgcctgaa g	aagcgtcat acggaccg	ga ttgttttcgc 60
tggcccagtg tcc	ccggagc ttgtgtgcga t	acagagagc acctcgga	ag ctgaggcagc 120
tggtacttga cag		tcg acc ata gtc tc Ser Thr Ile Val Se 5	
Gln Ile Lys Ar	gg aag gct ccc cgt gg gg Lys Ala Pro Arg Gl	gc ttt cta aag cga Ly Phe Leu Lys Arg	gtc ttc aag 219
	ct caa ctt cgt ctg ga ro Gln Leu Arg Leu Gl 35		

cat ctg aac tgt tta ctg ttt gtt cat cga tta gca gaa gag tcc agg

His Leu Asn Cys Leu Leu Phe Val His Arg Leu Ala Glu Glu Ser Arg

50

45

ca aa	c gct	tgt	gcg	agt	aaa	tgt	aga	gtc	att	aac	aag	gag	cat	gta		363	)
hr As	n Ala	Cys	Ala	Ser	Lys	Çys	Arg	Val	Ile	Asn	Lys	Glu	His	Val			
60				65					70					75	)	•	
etg go	c gca	gca	aag	gta	att	cta	aag	aag	agc	aga	ggt		ta	gaag	tc	410	)
Leu Al	a Ala	Ala	Lys	Val	Ile	Leu	Lys	Lys	Ser	Arg	Gly						
			80					85									
aagaa	cata	ttct	tgaa	ag t	tatg	atgc	a tt	cttt	tggg	tggt	taaca	aga	tcat	aaag	ac	470	
atttt	taca	catca	agtt	aa t	atgg	gatt	a tt	aaat	attg	gata	ataaa	aaa	aaaa	aaaa	laa	530	1
3.								•								533	Į
	•																
-010\																	
<210>	4																
<211>	88																

<400> 4

<212> PRT

<213> Homo sapiens

Met Ala Leu Ser Thr Ile Val Ser Gln Arg Lys Gln Ile Lys Arg Lys

1 5 10 15

Ala Pro Arg Gly Phe Leu Lys Arg Val Phe Lys Leu Lys Lys Pro Gln
20 25 30

Leu Arg Leu Glu Lys Ser Gly Asp Leu Leu Val His Leu Asn Cys Leu 35 40 45

Leu Phe Val His Arg Leu Ala Glu Glu Ser Arg Thr Asn Ala Cys Ala 50 55 60

Ser Lys Cys Arg Val Ile Asn Lys Glu His Val Leu Ala Ala Ala Lys 65 70 75 80

Val Ile Leu Lys Lys Ser Arg Gly

85

<210	>	5														
<211	>	206	7													
<212	>	DNA														
<213	>	Hom	o sa	pien	s											
<220	>															
<221	>	CDS	i													
<222	>	(32	) (	1384	)											
<223	>	Clo	ne G	E6												
<400	>	5														
cccg	ggct	ge c	aggo	gccc	a go	tgtg	icccs	ιg			atg	gat	ggg	aca	gag	46
_		-		-	_			_			Met	Asp	Gly	Thr	Glu	
											1				5	
acc	cgg	cag	cgg	agg	ctg	gac	agc	tgt	ggc	aag	cca	ggg	gag	ctg	ggg	94
Thr	Arg	Gln	Arg	Arg	Leu	Asp	Ser	Cys	Gly	Lys	Pro	Gly	Glu	Leu	Gly	
				10					15		,			20		. •
				ctc	-											142
Leu	Pro	His	Pro	Leu	Ser	Thr	Gly	Gly	Leu	Pro	Val	Ala		Glu	Asp	
			25					30					35			
					_											
	-			gcc												
СТА	ATA		Arg	Ala	Pro	GIU		GIN	ser	vai	Int	50	гуу	PIO	ьец	
		40					45					30				
σασ	act	σασ	cct	agc	agg	αaσ	acc	acc	taa	tcc	ata	ggc	ctt	caq	ata	238
				Ser												
	55				_	60			•		65					•
		•	•	•												
acc	gtg	ccc	ttc	atg	ttt	gca	ggc	ctg	gga	ctg	tcc	tgg	gcc	ggc	atg	286
Thr	Val	Pro	Phe	Met	Phe	Ala	Gly	Leu	Gly	Leu	Ser	Trp	Ala	Gly	Met	
70					75					80					85	
				ttc												
Leu	Leu	Asp	Tyr	Phe	Gln	Ala	Asn	Thr	Gly	Gln	Ile	Asp	Asp			ı
				90					95					100		

	-		-	-						gcc						382
Glu	Gln	His	Arg	Val	Ile	Ser	Ser	Asn-	Leu	Ala	Leu	Ile	Gln	Val	Gln	
			105					110					115			
_										gct						430
Ala	Thr		Val	Gly	Leu	Leu		Ala	Val	Ala	Ala		Leu	Leu	CTÀ.	
		120					125					130				
		<b>.</b>							~~~		~+~	~~~	++~	ata	tat	478
										aag Lys						470
vaı		ser	Arg	GIU	GIU	140	Asp	Val	Ата	гуз	145	GIU	neu	Беи	Cys	
	135					140					143					
acc	age	aαt	atc	ctc	act	acc	ttc	ctt	σca	gcc	ttt	acc	cta	aga	ata	526
-	-	_	-							Ala						
150					155					160				•	165	
ctg	atg	gtc	tgt	ata	gtg	att	ggt	gct	cga	aag	ctc	ggg	gtc	aac	cca	574
Leu	Met	Val	Cys	Ile	Val	Ile	Gly	Ala	Arg	Lys	Leu	Gly	Val	Asn	Pro	
				170					175					180		
			,								•					
gac	aac	att	gcc	acg	ccc	att	gca	gcc	ago	ctg	gga	gac	ctc	ato	aca	622
Asp	Asn	Ile	Ala	Thr	Pro	Ile	Ala	Ala	Ser	Leu	Gly	Asp	Leu	Ile	Thr	
			185					190					195			
															gat	670
Leu	Ser	Ile	Leu	Ala	Leu	Val	Ser	Ser	Phe	Phe	Tyr	Arg	His	Lys	Asp	
		200	1				205	,				210	}			
															acc	718
Ser	Arg	Туг	Leu	Thr	Pro			. Cys	Let	ı Ser			Ala	Let	Thr	
	215	•				220	)				225	)				
													. ~+-		ato	766
															g atc s Ile	700
		LIL	y vas	. rec			. пус	, GII	1 56	240		, 110	s va.	L LJ.	245	
230	,				235	,				230	,				2.0	
cto	ı aar	1 <b>+</b> +1	. ממי	tar	r tto	cea	a atr	at/	a ch	a acc	ato	ı atı	c ato	a a o	c agt	814
															r Ser	<b>7-1</b>
200	y	111	- 9-1	250					25					26		
				200	-					_						

							aaa Lys									862	
	<b></b> 3	<b></b> ,	265				•	270				-	275			·	
	-	-					gtc									910	
Gly	Met	Ala 280	Ile	Phe	Thr	Pro	Val 285	IIe	Cys	GTÀ	vaı	290	GTĀ	ASN	ъеп		
							atc									958	
Val	Ala 295	Ile	Gln	Thr	Ser	Arg 300	Ile	Ser	Thr	Tyr	Leu 305	His	Met	Trp	Ser		
							cag									1006	
Ala 310	Pro	Gly	Val	Leu	Pro 315	Leu	Gln	Met	Lys	Lys 320	Phe	Trp	Pro	Asn	Pro 325		
-															gtc	1054	
Cys	Ser	Thr	Phe	330		Ser	Glu	Ile	Asn 335	Ser	Met	Ser	· Ala	Arg 340	Val		
															atc	1102	?
Leu	Leu	Lev	345		. Val	Pro	Gly	His 350		Ile	Phe	Phe	355		e Ile		
															gtg	1150	)
Tyr	Let	360		ı Gly	y Glr	sei	365		e Asn	Se:	: Glī	37(		e Val	L Val		
															c ctg	1198	В
Let	37:		ı Le	u. Ala	a Gly	у Let 38		e Glr	n Val	LTh	385		u Lei	д Ту:	r Leu		
															t gac		6
Ala 39		u Va	l Me	t Va	1 Ar		u Th	r Tr	p His	40		a Le	u As	p Pr	o Asp 405		
															t act		4
As	n Hi	s Cy	s Il	e Pr 41		r Le	u Th	r Gl	y Le:		y As	p Le	u Le	u Gl 42	y Thr :0		

Len																
	Leu	Ala	Leu	Cys	Phe	Ьµ́е	Thr	Asp	Trp	Leu	Leu	Lys	Ser	Lys		
		425					430					435				
gag	ctg	ggt	ggc	atc	tca	gaa	ctg	gca	tct	gga	cct	ccc		taact	tg	1390
Glu	Leu	Gly	Gly	Ile	Ser	Glu	Leu	Ala	Ser	Gly	Pro	Pro				
	440					445					450					
ccg	ctg	gtcc	catt	tg c	tcati	tagaa	a tt	tcct	ctca	cat	cagt	ggg	atac	agaat	tt	1450
tttc	tcc	cttg	ccag	gt c	cttg	ggat	ggt	tgac	ccct	gcc.	tctg	cag	tagc	cttt	tg	1510
gtct	gct	aagg	tage	tc t	caca	cacci	t cg	gctc	tggg	gtt	gata	cct	gagc	ctgc	aa	1570
												•				1.620
agcc	ctg	aaat	caag	ag c	atgg	cttg	a gt	gtgt	gaat	atg	atgt	gtg	caca	tgct	ta	1630
																1690
agcg	tgc	aagt	gtgc	ac a	cgtt	tgtg	g ag	agga	gggc	gtt	cugg	CCL	yaya	aggc	aa	1090
						~~~~	~ ~+	~+~+	++~~	+	+++	cat	acco	estac	22	1750
agag	gca	tgtc	cagt	at g	CLLL	gcag	y y c	gtgt	ctyc	CCC		cat	gccc	acyc	au	1,30
2025	+~~	~~+~	~~~			aact	c tt	++c+	atto	CCA	acco	tca	gaac	rtett	ga	1810
ayaı	Lyg	ggtg	yayc	ay y	aayy	agcc			geco	CCu	ugu		guut	,,,,,,	-9-	2020
ataa	ctt	actt	acta	ıtc t	tcac	cagg	t to	:aagc	teca	t.aa	acca	cac	tact	acta	ta	1870
		accc	gorg	,		0499			9	-99	,		-5-	- 5 5	3	
agaa	aat	gtac	agco	tc c	ccaq	gatg	a ac	rcctc	atac	aac	cctt	cat	ctg	cacto	aa	1930
		,	,			J J	, ,,						_			
ttaa	tca	tato	ctto	ict c	tctt	ttta	t tt	tect	tttt	gtt	tgtt	agc	aaaa	aacct	ct	1990
	9			,	,					-		-				
taga	ttt	caat	aato	ag a	agaag	tgta	a aa	ataaa	acag	att	atat	tgt	aaaa	aaaaa	ıaa	2050
-									_			-				
aaaa	iaaa	aaaa	aaa													2067
	Glu cccgg	Glu Leu 440  cccgctg  cttctcc  gtctgct  agccctg  agcgtgc  agaggca  agattgg  gtggctt  agaaggt  ttaatcg  tagattt	gag ctg ggt Glu Leu Gly 440  cccgctg gtcc  ttctcc cttg gtctgct aagg agccctg aaat agcgtgc aagt agaggca tgtc agattgg ggtg gtggctt actt agaaggt gtac ttaatcg tgtc tagattt caat	gag ctg ggt ggc Glu Leu Gly Gly 440  cccgctg gtcccatt  cttctcc cttgccag gtctgct aaggtagc agccctg aaatcaag agcgtgc aagtgtgc agaggca tgtccagt agattgg ggtggagc gtggctt acttgctg agaaggt gtacagcc ttaatcg tgtccttg	gag ctg ggt ggc atc Glu Leu Gly Gly Ile 440  cccgctg gtcccatttg c  ttctcc cttgccaggt c  gtctgct aaggtagctc t  agccctg aaatcaagag c  agaggca tgtccagtat g  agattgg ggtggagcag g  gtggctt acttgctgtc t  agaaggt gtacagcctc c  ttaatcg tgtccttgct g  tagattt caataatcag a	gag ctg ggt ggc atc tca Glu Leu Gly Gly Ile Ser 440  cccgctg gtcccatttg ctcatt  cttctcc cttgccaggt ccttgc  gtctgct aaggtagctc tcacac  agccctg aaatcaagag catggc  agcgtgc aagtgtgcac acgtt  agaggca tgtccagtat gcttt  agattgg ggtggagcag gaagg  gtggctt acttgctgtc ttcacc  agaaggt gtacagctc cccag  ttaatcg tgtccttgct gtctt  tagattt caataatcag agaag	gag ctg ggt ggc atc tca gaa Glu Leu Gly Gly Ile Ser Glu 440 445 cccgctg gtcccatttg ctcattagaa cttctcc cttgccaggt ccttgggate gtctgct aaggtagctc tcacacacce agccctg aaatcaagag catggcttga agcgtgc aagtgtgcac acgtttgtga agaggca tgtccagtat gctttgcag agattgg ggtggagcag gaaggagct gtggctt acttgctgtc ttcaccagg agaaggt gtacagcctc cccaggatg ttaatcg tgtccttgct gtctttta tagattt caataatcag agaagtgta	gag ctg ggt ggc atc tca gaa ctg Glu Leu Gly Gly Ile Ser Glu Leu 440 445  cccgctg gtcccatttg ctcattagaa tt  cttctcc cttgccaggt ccttgggatg gt gtctgct aaggtagctc tcacacacct cg agccctg aaatcaagag catggcttga gt agcgtgc aagtgtgcac acgtttgtgg ag agaggca tgtccagtat gctttgcagg gt agattgg ggtggagcag gaaggagctc tt gtggctt acttgctgtc ttcaccaggt tc agaaggt gtacagcctc cccaggatgg gg ttaatcg tgtccttgct gtcttttat tt tagattt caataatcag agaagtgtaa aa	gag ctg ggt ggc atc tca gaa ctg gca Glu Leu Gly Gly Ile Ser Glu Leu Ala 440 445  cccgctg gtcccatttg ctcattagaa tttcctc cttctcc cttgccaggt ccttgggatg gttgac gtctgct aaggtagctc tcacacacct cggctc agccctg aaatcaagag catggcttga gtgtgt agcgtgc aagtgtgcac acgtttgtgg agagga agaggca tgtccagtat gctttgcagg gtgtgt agattgg ggtggagcag gaaggagctc ttttct gtggctt acttgctgtc ttcaccaggt tcaagc agaaggt gtacagcctc cccaggatgg ggcctc ttaatcg tgtccttgct gtcttttat tttcct tagattt caataatcag agaagtgtaa aataaa	gag ctg ggt ggc atc tca gaa ctg gca tct Glu Leu Gly Gly Ile Ser Glu Leu Ala Ser 440 445  cccgctg gtcccatttg ctcattagaa tttcctctca tttctcc cttgccaggt ccttgggatg gttgacccct gtctgct aaggtagctc tcacacacct cggctctggg agccctg aaatcaagag catggcttga gtgtgtgaat agcgtgc aagtgtgcac acgtttgtgg agaggaggt agaggca tgtccagtat gctttgcagg gtgtgtttgc agattgg ggtggagcag gaaggagctc ttttctgttc gtggctt acttgctgtc ttcaccaggt tcaagctccg agaaggt gtacagcctc cccaggatgg ggcctcatac ttaatcg tgtccttgct gtcttttat tttccttttt tagattt caataatcag agaagtgtaa aataaaacag	gag ctg ggt ggc atc tca gaa ctg gca tct gga Glu Leu Gly Gly Ile Ser Glu Leu Ala Ser Gly 440 445  cccgctg gtcccatttg ctcattagaa tttcctctca catc cttctcc cttgccaggt ccttgggatg gttgacccct gcc gtctgct aaggtagctc tcacacacct cggctctggg gtt agccctg aaatcaagag catggcttga gtgtgtgaat atg agcgtgc aagtgtgcac acgtttgtgg agaggagggt gtt agaggca tgtccagtat gctttgcagg gtgtgttgc tct agattgg ggtggagcag gaaggagctc ttttctgttc cca gtggctt acttgctgtc tcaccaggt tcaagctccg tgg agaaggt gtacagcctc cccaggatgg ggcctcatac aac ttaatcg tgtccttgct gtcttttat tttcctttt gtt tagattt caataatcag agaagtgtaa aataaaacag att	gag ctg ggt ggc atc tca gaa ctg gca tct gga cct Glu Leu Gly Gly Ile Ser Glu Leu Ala Ser Gly Pro 440 445 450  cccgctg gtcccatttg ctcattagaa tttcctctca catcagt  cttctcc cttgccaggt ccttgggatg gttgacccct gcctctg  gtctgct aaggtagctc tcacacacct cggctctggg gttgata  agccctg aaatcaagag catggcttga gtgtgtgaat atgatgt  agcgtgc aagtgtgcac acgtttgtgg agaggagggt gttctgg  agaaggca tgtccagtat gctttgcagg gtgtgtttgc tctttcc  agattgg ggtggagcag gaaggagctc tttctgttc ccaagcc  gtggctt acttgctgtc tcaccaggt tcaagctccg tgggcca  agaaggt gtacagcctc cccaggatgg ggcctcatac aaccctt  ttaatcg tgtccttgct gtcttttat tttccttttt gtttgtt  tagattt caataatcag agaagtgtaa aataaaacag attatat	gag ctg ggt ggc atc tca gaa ctg gca tct gga cct ccc Glu Leu Gly Gly Ile Ser Glu Leu Ala Ser Gly Pro Pro 440 445 450  cccgctg gtcccatttg ctcattagaa tttcctctca catcagtggg cttctcc cttgccaggt ccttgggatg gttgacccct gcctctgcag gtctgct aaggtagctc tcacacacct cggctctggg gttgatacct agccctg aaatcaagag catggcttga gtgtgtgaat atgatgtgt acgcgtgc aagtgtgcac acgtttgtgg agaggaggt gttctggcct agaggca tgtccagtat gctttgcagg gtgtgtttgc tctttccat agattgg ggtggagcag gaaggagctc ttttctgttc ccaagcctca gtggctt acttgctgtc ttcaccaggt tcaagctccg tgggccacac agaaggt gtacagcctc cccaggatgg ggcctcatac aacccttcat ttaatcg tgtccttgct gtcttttat tttcctttt gtttgttagc tagattt caataatcag agaagtgtaa aataaaacag attatattgt	gag ctg ggt ggc atc tca gaa ctg gca tct gga cct ccc Glu Leu Gly Gly Ile Ser Glu Leu Ala Ser Gly Pro Pro 440 445 450  cccgctg gtcccatttg ctcattagaa tttcctctca catcagtggg atac tttctcc cttgccaggt ccttgggatg gttgacccct gcctctgcag tagc gtctgct aaggtagctc tcacacacct cggctctggg gttgatacct gagc agccctg aaatcaagag catggcttga gtgtgtgaat atgatgtgtg caca agcgtgc aagtgtgcac acgtttgtgg agaggaggt gttctggcct gaga agaggca tgtccagtat gctttgcagg gtgtgtttgc tcttttccat gccc agattgg ggtggagcag gaaggagctc ttttctgttc ccaagcctca gaac gtggctt acttgctgtc ttcaccaggt tcaagctccg tgggccacac tgct agaaggt gtacagcctc cccaggatgg ggcctcatac aacccttcat ctgc ttaatcg tgtccttgct gtcttttat tttcctttt gtttgttagc aaac tagattt caataatcag agaagtgtaa aataaaacag attatattgt aaac tagattt caataatcag agaagtgtaa aataaaacag attatattgt aaac tagattt caataatcag agaagtgtaa aataaaacag attatattgt aaac	gag ctg ggt ggc atc tca gaa ctg gca tct gga cct ccc taacd Glu Leu Gly Gly Ile Ser Glu Leu Ala Ser Gly Pro Pro 440 445 450  cccgctg gtcccatttg ctcattagaa tttcctctca catcagtggg atacagaa cttctcc cttgccaggt ccttgggatg gttgacccct gcctctgcag tagccttt gtctgct aaggtagctc tcacacacct cggctctggg gttgatacct gagcctgc agccctg aaatcaagag catggcttga gtgtgtgaat atgatgtgtg cacatgct agcgtgc aagtgtgcac acgtttgtgg agaggaggt gttctggcct gagaaggt agaggca tgtccagtat gctttgcagg gtgtgtttgc tctttccat gcccatgc agagttgg ggtggagcag gaaggagctc ttttctgttc ccaagcctca gaactctt gtggctt acttgctgtc ttcaccaggt tcaagctccg tgggccacac tgctgctg agaaggt gtacagcctc cccaggatgg ggcctcatac aacccttcat ctgcactc ttaatcg tgtccttgct gtcttttat tttcctttt gtttgttagc aaaaacct tagattt caataatcag agaagtgtaa aataaaacag attatattgt aaaaaaacag	gag ctg ggt ggc atc tca gaa ctg gca tct gga cct ccc taactg Glu Leu Gly Gly Ile Ser Glu Leu Ala Ser Gly Pro Pro 440 445 450  cccgctg gtcccatttg ctcattagaa tttcctctca catcagtggg atacagaatt  cttctcc cttgccaggt ccttgggatg gttgacccct gcctctgcag tagcctttg gtctgct aaggtagctc tcacacacct cggctctggg gttgatacct gagcctgcaa  agccctg aaatcaagag catggcttga gtgtgtgaat atgatgtgt cacatgcta  agcgtgc aagtgtgcac acgtttgtgg agaggaggt gttctggcct gagaaggtaa  agaggca tgtccagtat gctttgcagg gtgtgtttgc tcttttccat gcccatgcaa  agagtgg ggtggagcag gaaggagctc tttcctgtc ccaagcctca gaactcttga  gtggctt acttgctgtc ttcaccaggt tcaagctccg tgggccacac tgctgctg  agaaggt gtacagcctc cccaggatgg ggcctcatac aacccttcat ctgcactcaa  ttaatcg tgtccttgct gtctttttat tttccttttt gtttgttagc aaaaacctct  tagattt caataatcag agaagtgtaa aataaaacag attatattgt aaaaaaaaaa

<210> 6

<211> 451

<212> PRT

<213> Homo sapiens

<400> 6

1et	Asp	стА	Thr	5	rnr	Arg	GIN	Arg .	10	neu	Asp	ser	cys	15	гÀЗ
Pro	Gly	Glu	Leu 20	Gly	Leu	Pro	His	Pro 25	Leu	Ser	Thr	Gly	<b>Gly</b> 30	Leu	Pro
Val	Ala	Ser 35	Glu	Asp	Gly	Ala	Leu 40	Arg	Ala	Pro	Glu	Ser 45	Gln	Ser	Val
Thr	Pro 50	Lys	Pro	Leu	Glu	Thr 55	Glu	Pro	Ser	Arg	Glu 60	Thr	Ala	Trp	Ser
Ile 65	Gly	Leu	Gln	Val	Thr 70	Val	Pro	Phe	Met	Phe 75	Ala	Gly	Leu	Gly	Leu 80
Ser	Trp	Ala	Gly	Met 85	Leu	Leu	Asp	Tyr	Phe 90	Gln	Ala	Asn	Thr	Gly 95	Gln
Ile	Asp	Asp	Pro 100		Glu	Gln	His	Arg 105	Val	Ile	Ser	Ser	Asn 110	Leu	Ala
Leu	Ile	Gln 115		Gln	Ala	Thr	Val 120	Val	Gly	Leu	Leu	Ala 125	Ala	Val	Ala
Ala	Leu 130		Leu	Gly	Val	Val 135	Ser	Arg	Glu	Glu	Val		Val	Ala	Lys
Val 145		Leu	Leu	_	Ala . 150		Ser	Val	Leu	155		Phe	. Leu	Ala	Ala 160
Phe	Ala	Leu	Gly	v Val . 165		Met		Cys	Ile 170		. Ile	e Gly	Ala	Arg 175	Lys
Leu	Gly	va]	Asr 180		Asp	Asn	Ile	Ala 185		Pro	) Ile	e Ala	190		Leu
Gly	/ Asp	Le:		e Thr	Leu	ı Ser	200		ı Ala	a Let	ı Val	L Sei 20!		. Phe	. Phe
Туг	. Arc	, Hi:	s Lv:	s Ast	Ser	. Arc	туг	Let	ı Thı	r Pro	o Let	ı Vai	l Cys	s Leu	Ser

220

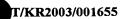
	210					215					220				
Phe 225	Ala	Ala	Leu	Thr	Pro 230	Val	Trp	Val	Leu	Ile 235	Ala ·	Lys	Gln		Pro 240
Pro	Ile	Val	Lys	Ile 245	Leu	Lys	Phe	Gly	Trp 250	Phe	Pro	Ile	Ile	Leu 255	Ala
Met	Val	Ile	Ser 260	Ser	Phe	Gly	Gly	Leu 265	Ile	Leu	Ser	Lys	Thr 270	Val	Ser
Lys	Gln	Gln 275	Tyr	Lys	Gly	Met	Ala 280	Ile	Phe	Thr	Pro	Val 285	Ile	Cys	Gly
Val	Gly 290	Gly	Asn	Leu	Val	Ala 295	Ile	Gln	Thr	Ser	Arg 300	Ile	Ser	Thr	Tyr
Leu 305	His	Met	Trp	Ser	Ala 310		Gly	Val	Leu	Pro 315		Gln	Met	Lys	Lys 320
Phe	Trp	Pro	Asn	Pro 325		Ser	Thr	Phe	Cys 330		Ser	Glu	Ile	Asn 335	Ser
Met	Ser	Ala	Arg 340		Leu	Leu	Leu	Leu 345		Val	Pro	Gly	His 350		Ile
Phe	Phe	Туг 355		: Ile	туг	Leu	Val 360		Gly	Gln	. Ser	7al		Asn	Ser
Gln	370		. Val	. Val	. Leu	ту: 375		Let	ı Ala	Gly	, Leu 380		e Gln	Val	Thr
Ile 385		ı Lev	ту:	Let	390		ı Val	. Met	: Val	39!		t Thi	r Trp	His	Gln 400
Ala	ı Leı	ı Ası	Pro	Ası 409		n Hi	s Cys	s Ile	e Pro		r Let	ı Th:	r Gly	/ Let	ı Gly
Asp	) Le	ı Lei	420	_	r Gl	y Le	u Lei	u Ala		u Cy	s Phe	e Ph	e Thi		Trp

Leu Leu Lys Ser Lys Ala Glu Leu Gly Gly Ile Ser Glu Leu Ala Ser 445 440 Gly Pro Pro 450 <210> <211> 2178 <212> DNA <213> Homo sapiens <220> <221> CDS <222> (41)..(1501) <223> Clone MB5 <400> 7 gggaggggac ccgggctgcc aggcgcccag ctgtgcccag atg gat ggg aca gag 55 Met Asp Gly Thr Glu 1 103 acc cgg cag cgg agg ctg gac agc tgt ggc aag cca ggg gag ctg ggg Thr Arg Gln Arg Arg Leu Asp Ser Cys Gly Lys Pro Gly Glu Leu Gly 15 20 10 151 ctt cct cac ccc ctc agc aca gga gga ctc cct gta gcc tca gaa gat Leu Pro His Pro Leu Ser Thr Gly Gly Leu Pro Val Ala Ser Glu Asp 25 gga gct ctc agg gcc cct gag agc caa agc gtg acc ccc aag cca ctg 199 Gly Ala Leu Arg Ala Pro Glu Ser Gln Ser Val Thr Pro Lys Pro Leu 40 gag act gag cct agc agg gag acc gcc tgg tcc ata ggc ctt cag gtg 247 Glu Thr Glu Pro Ser Arg Glu Thr Ala Trp Ser Ile Gly Leu Gln Val 65 55 60

				atg													295
Thr	Val	Pro	Phe	Met	Phe	Ala	Gly	Leu	Gly	Leu	Ser	Trp	Ala	Gly	M	et	
70					75					80					1	85	
	_			ttc													343
Leu	Leu	Asp	Tyr	Phe	Gln	His	Trp	Pro	Val	Phe	Val	Glu	Val	ГÀз	A	.sp	
				90					95					100	)		
			•														
				gtg													391
Leu	Leu	Thr	Leu	Val	Pro	Pro	Leu	Val	Gly	Leu	ГÀЗ	Gly	Asn	Let	1 G	lu	
			105					110					115				
																	400
				tcc													439
Met	Thr	Leu	Ala	Ser	Arg	Leu	Ser	Thr	Ala	Ala				Gli	n, I	lle	
		120					125				•	130	)				
																	407
				gag													487
Asp	Asp	Pro	Gln	Glu	Gln	His	Arg	Val	. Ile	Ser			ı Leı	ı Al	a 1	Leu	
	135	<b>,</b>				140	١ .				145	•					
				•													525
				g gcc													535
Ile	Glr	val	l Glr	ı Ala	Thi	c Val	. Val	L Gl	Leu			a Ala	a Va	I AI			
150	)				155	5				160	)					165	
																	503
				c gto													583
Le	ı Leı	ı Le	u Gl	y Vai	l Va	l Se	r Ar	g Gl			l Ası	p Va	1 Al			Val	
				170	0				17	5				18	30		
										_							631
				t gc													631
G1	u Le	u Le	u Cy	s Al	a Se	r Se	r Va			r Al	a Ph	е ге			La	Pne	
			18	5				19	0				. 19	15			
			•														670
				g ct													679
Al	a Le			l Le	u Me	t Va			e Va	1 11	e GI			g L	ys	ьеи	
		20	0				20	5				21	·U				
											<b>.</b>				+~	~~~	.727
				a ga													121
G1			n Pr	o As	p As			a Th	ır Pr	.Ó 11			La 31	- L L	eu	GTĀ	
	21	.5				22	0			•	22	2.5					

-				ctg Leu													775
230					235					240					245		
-			-	agt													823
Arg	HIS	ъàs	Asp	Ser 250	Arg	IÄL	Dea	1111	255	лец	Vai	Cys	Бец	260	1110		
				cca													871
Ala	Ala	Leu	Thr 265	Pro	Val	Trp	Val	Leu 270	Ile	Ala	Lys	Gln	Ser 275	Pro	Pro		
				ctg													919
Ile	Val	Lys 280	Ile	Leu	Lys	Phe	Gly 285	Trp	Phe	Pro	Ile	11e 290	Leu	Ala	Met		
															aaa		967
Val	11e 295		Ser	Phe	Gly	300		Ile	Leu	Ser	Lys 305		Val	Ser	Lys		
															gtt		1015
Gln 310		Туг	Lys	Gly	Met 315		Ile	Phe	Thr	9ro 320		Ile	: Cys	G17	7 Val 325		
															ctg		1063
Gly	. G1 <sup>7</sup>	/ Asr	. Let	1 Val 330		l Ile	: Gln	Thr	335		, Ile	e Ser	Thr	34(	Leu		
															a ttc		1111
His	Met	Trp	345		e Pro	o Gly	/ Val	350		) Let	ı Glr	ı Met	355		s Phe	•	
															c atg		1159
Trp	Pro	36		o Cys	s Se:	r Thi	36		5 Thi	c Se:	r Glu	37		n Se	r Met		
															t ttc		1207
Sei	37		g Va	l Le	n Fe	u Lei 38		u Vai	l Va	l Pr	o Gl; 38		s Le	u Il	e Phe		
	- '	-															

ttc	tac	atc	atc	tac	ctg	gtg	gag	ggt	cag	tca	gtc	ata	aac	agc	cag	1255
Phe	Tyr	Ile	Ile	Tyr	Leu	Val	Glu	Gly	Gln	Ser	Val	Ile	Asn	Ser	Gln	
390					395					400					405	
acc	ttt	gtg	gtg	ctc	tac	ctg	ctg	gca	ggc	ctg	atc	cag	gtg	aca	atc	1303
Thr	Phe	Val	Val	Leu	Tyr	Leu	Leu	Ala	Gly	Leu	Ile	Gln	Val	Thr	Ile	
				410					415					420		
ctg	ctġ	tac	ctg	gca	gaa	gtg	atg	gtt	cgg	ctg	act	tgg	cac	cag	gcc	1351
				Ala												
		-	425					430					435			
cto	gat	cet	gac	aac	cac	tac	atc	ccc	tac	ctt	aca	ggg	ctg	ggg	gac	1399
				Asn												
Бес	. Ast	440				-7-	445		-3			450		-	_	
		, 440														
		. ~~+			. ctc	cta	~~~	cto	tac	. +++	tto	: act	gac	: tac	g cta	1447
-															Leu	-
rei			rni	: Слу	пес			. Dec	. Cys		465					
•	45	5				460					400	,				
									+.	. +a-		cto		. +~1	t aaa	1495
															t gga	1433
		s Sei	c Lys	s Ala			ı Gış	, GTZ	, 116			ı be	ı mı	1 56.	r Gly	
47	)				475	•				480	,				485	
																1550
CC	t cc	С	taa	actg	ggc (	ccg	etggi	tc co	catti	tgcto	c at1	taga	attt	cct.	ctcacat	1550
Pr	o Pr	0				-										
ca	gtgg	gata	cag	aatt	cag	tttc	tccc	tt g	ccag	gtcc	t tg	ggat	ggtt	gac	ccctgcc	1610
tc	tgca	gtag	cct	tttg	tga ·	gtct	gcta	ag g	tage	tctc	a ca	cacc	tcgg	ctc	tggggtt	1670
																·.
ga	tacc	tgag	cct	gcaa	tag	agcc	ctga	aa t	caag	agca	t gg	cttg	agtg	tgt	gaátato	1730
							:									
at	gtgt	gcac	atg	ctta	atg	agcg	tgca	ag t	gtgc	acac	g tt	tgtg	gaga	gga	ıgggtgtt	1790
ct	ggc	tgag	aag	gtaa	aga	agag	gcat	gt c	cagt	atgc	t tt	gcag	ggtg	tgt	ttgctc	1850
tt	tcca	atgco	cat	gcaa	ccc	agat	tggg	gt g	gago	agga	a gg	agct	cttt	: tct	gttccc	a 1910
		~														
aç	ject	cagaa	cto	ttga	ıgct	gtgg	ctta	ct t	gctg	tctt	c ac	cago	gttca	a ago	ctccgtg	g 1970
_		_														



gccacactgc tgctgtgcca agaaggtgta cagcctcccc aggatgggc ctcatacaa	c 2030
ccttcatctg cactcaacat ttaatcgtgt ccttgctgtc tttttatttt cctttttgt	t 2090
tgttagcaaa aacctctatt tagatttcaa taatcagaga agtgtaaaat aaaacagat	t 2150
atattgtact tgaaaaaaaa aaaaaaaa	2178
<210> 8	
•	
<211> 487	
<212> PRT	
<213> Homo sapiens	
<400> 8	
Met Asp Gly Thr Glu Thr Arg Gln Arg Arg Leu Asp Ser Cys Gly Lys	
1 5 10 15	
Pro Gly Glu Leu Gly Leu Pro His Pro Leu Ser Thr Gly Gly Leu Pro	
20 25 30	•
Val Ala Ser Glu Asp Gly Ala Leu Arg Ala Pro Glu Ser Gln Ser Val	
35 40 45	
Thr Pro Lys Pro Leu Glu Thr Glu Pro Ser Arg Glu Thr Ala Trp Ser	
-	
50 55 60	
Ile Gly Leu Gln Val Thr Val Pro Phe Met Phe Ala Gly Leu Gly Leu	
65 70 75 80	
Ser Trp Ala Gly Met Leu Leu Asp Tyr Phe Gln His Trp Pro Val Phe	
85 90 95	
Val Glu Val Lys Asp Leu Leu Thr Leu Val Pro Pro Leu Val Gly Leu	
100 105 110	
Lys Gly Asn Leu Glu Met Thr Leu Ala Ser Arg Leu Ser Thr Ala Ala	
115 120 125	

	130	СΙΆ	GIN	TTE	-	135	PIO	GIII	GIU	GIII	140	wig	Val	TTC	Ser
Ser 145	Asn	Leu	Ala	Leu	Ile 150	Gln	Val	Gln	Ala	Thr 155	Val	Val	Gly		Leu 160
Ala	Ala	Val	Ala	Ala 165	Leu	Leu	Leu	Gly	Val 170	Val	Ser	Arg	Glu	Glu 175	Val
Asp	Val	Ala	Lys 180	Val	Glu	Leu	Leu	Cys 185	Ala	Ser	Ser	Val	Leu 190	Thr	Ala
Phe	Leu	Ala 195	Ala	Phe	Ala	Leu	Gly 200	Val	Leu	Met	Val	Cys 205	Ile	Val	Ile
Gly	Ala 210	Arg	Lys	Leu	Gly	Val 215	Asn	Pro	Asp	Asn	11e 220	Ala	Thr	Pro	Ile
Ala 225	Ala	Ser	Leu	Gly	Asp 230	Leu	Ile	Thr	Leu	Ser 235		Leu	Ala	Leu	Val 240
Ser	Ser	Phe	Phe	Tyr 245	Arg	His	Lys	Asp	Ser 250		Tyr	Leu	Thr	Pro 255	Leu
Val	Cys	Leu	Ser 260		Ala	Ala	Leu	Thr 265		Val	Trp	Val	Leu 270	Ile	Ala
Lys	Gln	Ser 275		Pro	Ile	Val	Lys 280		e Lev	ı Lys	. Phe	Gly 285		Phe	Pro
Ile	290		ı Ala	Met	Val	11e		Ser	: Phe	e Gl <sup>'</sup>	, Gly 300		ı Ile	Leu	Ser
Lys		· Val	. Ser	. Lys	310		туг	Lys	s Gly	y Met 315		ı Ile	Phe	Thr	Pro 320
Va]	. Ile	e Cys	s Gly	7 Val		, Gl	Y Ası	1 Let	ı Va:		a Ile	e Glr	1 Thr	335	Arg
Ile	e Sei	Th	r Tyı	Lei	ı His	s Met	t Tr	o Se:	r Ala	a Pr	o G1	y Val	l Leu	Pro	Leu

340 345 350

Gln Met Lys Lys Phe Trp Pro Asn Pro Cys Ser Thr Phe Cys Thr Ser . 355 360 365

Glu Ile Asn Ser Met Ser Ala Arg Val Leu Leu Leu Leu Val Val Pro 370 375 380

Gly His Leu Ile Phe Phe Tyr Ile Ile Tyr Leu Val Glu Gly Gln Ser 385 390 395 400

Val Ile Asn Ser Gln Thr Phe Val Val Leu Tyr Leu Leu Ala Gly Leu 405 410 415

Ile Gln Val Thr Ile Leu Leu Tyr Leu Ala Glu Val Met Val Arg Leu 420 425 430

Thr Trp His Gln Ala Leu Asp Pro Asp Asn His Cys Ile Pro Tyr Leu
435 440 445

Thr Gly Leu Gly Asp Leu Leu Gly Thr Gly Leu Leu Ala Leu Cys Phe 450 455 460

Phe Thr Asp Trp Leu Leu Lys Ser Lys Ala Glu Leu Gly Gly Ile Ser 465 470 475 480

Glu Leu Ala Ser Gly Pro Pro 485

<210> 9

<211> . 1616

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (31)..(1551)

<223> Clone IE4

<400	)>	9																
ccg	gctg	rcc a	ggcg	ccca	g ct	gtgc	ccag	atg	gat	ggg	aca	gag	acc	cgg	cag			54
								Met	Asp	Gly	Thr	Glu	Thr	Arg	Gln			
								1	•			5	•					
	agg				<b></b>	~~~		007	~~~	~~~	cta	aaa	c++	cct	CPC			102
	Arg																	102
ALG	10	Den	nop	DGL	Cys	15	Lys		C±3	024	20	OJ						
														-				
ccc	ctc	agc	aca	gga	gga	ctc	cct	gta	gcc	tca	gaa	gat	gga	gct	ctc			150
Pro	Leu	Ser	Thr	Gly	Gly	Leu	Pro	Val	Ala	Ser	Glu	Asp	Gly	Ala	Leu			
25					30					35					40			
	gcc																	198
Arg	Ala	Pro	Glu	Ser	Gln	Ser	Val	Thr		Lys	Pro	Leu	Glu		Glu			
				45					50					55				
						•												246
	agc				-													246
PIO	Ser	Arg	60	THE	Ата	тър	Ser	65	GIŞ	Den	GTII	Val	70	Val	FIO	•		•
			00					03					, ,					
tto	atg	ttt	gca	ggc	ctg	gga	ctg	tcc	tgg	gcc	ggc	atg	ctt	ctg	gac			294
	Met																	
		75	,				80					85						
tat	ttc	cag	cac	tgg	cct	gtg	ttt	gtg	gag	gtg	aaa	gac	ctt	ttg	aca			342
туг	Phe	Gln	His	Trp	Pro	Val	Phe	Val	Glu	Val	Lys	Asp	Leu	Leu	Thr			
	90					95					100							
	gtg													-			•	390
	ı Val	. Pro	Pro	Leu		_	Leu	Lys	GTA			GLu	Met	Thr	_ ьец 120			
105	•				110					115					120			
αca	tec	. aga	cto	tcc	aca	act	acc	aac	act	gga	caa	att	gat	gac	ccc			438
-		_													Pro			
		:		125					130				-	135			•	
ca	g gag	j ca	g cac	aga:	gto	ato	ago	ago	aac	ctg	gco	cto	ato	cag	gtg			486
Gl	n Glu	ı Glı	n His	Arg	val	. Ile	Ser	Ser	Asn	Lev	Ala	Lev	ı Ile	Glr	Val			

			140					145					150				
cag	gcc	act	gtc	gtg	ggg	ctc	ttg	gct	gct	gtg	gct	gcg	ctg	ctg	ttg		534
Gln	Ala	Thr	Val	Val	Gly	Leu	Leu	Ala	Ala	Val	Ala	Ala	Leu	Leu	Leu		•
		155					160		٠		٠	165					
ggc	gtg	gtg	tct	cga	gag	gaa	gtg	gat	gtc	gcc	aag	gtg	gag	ttg	ctg		582
Gly	Val	Val	Ser	Arg	Glu	Glu	.Val	Asp	Val	Ala	Lys	Val	Glu	Leu	Leu		
	170					175					180						
tgt	gcc	agc	agt	gtc	ctc	act	gcc	ttc	ctt	gca	gcc	ttt	gcc	ctg	ggg		630
Cys	Ala	Ser	Ser	Val	Leu	Thr	Ala	Phe	Leu	Ala	Ala	Phe	Ala	Leu	Gly		
185					190					195					200		
	7	_	-	_					_	_	_		ggg				678
Val	Leu	Met	Val	Cys	Ile	Val	Ile	Gly	Ala	Arg	Lys	Leu	Gly	Val	Asn		
				205					210					215			
cca	gac	aac	att	gcc	acg	ccc	att	gca	gcc	agc	ctg	gga	gac	ctc	atc		726
Pro	Asp	Asn	Ile	Ala	Thr	Pro	Ile	Ala	Ala	Ser	Leu	Gly	Asp	Leu	Ile		
•			220					225					230				
aca	ctg	tcc	att	ctg	gct	ttg	gtt	agc	agc	ttc	ttc	tac	aga	cac	aaa		774
Thr	Leu	Ser	Ile	Leu	Ala	Leu	Val	Ser	Ser	Phe	Phe	Tyr	·Arg	His	Lys		
		235					240					245					
gat	agt	cgg	tat	ctg	acg	ccg	ctg	gtc	tgc	ctc	agc	ttt	gcg	gct	ctg		822
Asp	Ser	Arg	Tyr	Leu	Thr	Pro	Leu	Val	Cys	Leu	Ser	Phe	Ala	Ala	Leu		
	250	1				255					260						
acc	cca	gtg	tgg	gto	cto	att	gcc	aag	cag	ago	cca	ccc	atc	gtg	aag		870
Thr	Pro	Val	Trp	Val	. Leu	Ile	Ala	Lys	Gln	Ser	Pro	Pro	Ile	Val	Lys		
265					270					275	•				280	٠	
ato	cto	aaç	, ttt	ggc	: tgg	tto	сса	ato	ato	ctg	gcc	atg	gto	ato	agc		918
Ile	Let	Lys	Phe	e Gly	, Trp	Phe	Pro	Ile	: Ile	Leu	ı Ala	Met	. Val	Ile	Ser		
				285	<b>ò</b>				290	)				295	5		
agt	tto	gga	a gga	cto	ato	: ttg	ago	aaa	acc	gtt	tct	. aaa	a cag	cag	, tac		966
Ser	Phe	e Gly	/ Gly	Let	ılle	Lev	Ser	Lys	Thi	· Val	. Ser	Lys	Gln	Glr	ı Tyr		

			300					305					310			
				ata												1014
ŗ	GLY		Ата	Ile	Phe	Thr	320	vaı	TTE	Cys	GIÀ	325	GŢĀ	СТА	ASI	
		315					320					323				•
ctg	gtġ	gcc	att	cag	acc	agc	cga	atc	tca	acc	tac	ctg	cac	atg	tgg	1062
				Gln												
	330					335					340					
				gtc												1110
Ser	Ala	Pro	Gly	Val	Leu	Pro	Leu	Gln	Met		Lys	Phe	Trp	Pro		
345					350					355					360	
	h			ttc	+~~	200	tas	~==	ato	22+	tcc	ata	tca	act	caa	1158
				Phe												1130
FIO	Cys	Jer	* ***	365	Cys	11	001	0	370					375	5	
gtc	ctg	ctc	ttg	ctg	gtg	gtc	cca	ggc	cat	ctg	att	ttc	ttc	tac	atc	1206
Val	Leu	Leu	Leu	Leu	Val	Val	Pro	Gly	His	Leu	Ile	Phe	Phe	Tyr	Ile	
			380					385					390			
																•
				gag												1254
Ile	Tyr	Leu	Val	Glu	Gly	Gln	Ser	Val	Ile	Asn	Ser			Phe	Val	
		395					400					405				
							- •									1202
				ctg												1302
Val	410		ьеи	Leu	HIA	415		116	GIII	Val	420		Дea	пеа	ry-	
	310					710										
ctq	gca	gaa	gtq	atg	gtt	cqq	ctg	act	tgg	cac	cag	gco	ctg	gat	cct	1350
															Pro	
425					430					435		•		•	440	
gac	aac	cac	: tgc	ato	ccc	: tac	ctt	aca	ggg	ctg	ggg	gad	cto	cto	ggt	1398
Asp	Asn	His	Сув	; Ile	Pro	туг	Leu	Thr	Gly	Leu	Gly	/ Asp	Lev	ı Lev	Gly	
				445	•				450	)				455	i	
																1446
															gcc	1446
ser	Ser	: Sei	· val	r GTA	и на	a nur	: ATS	LAL	ı val	LIC	, AIG	ALC	, cys	Ini	: Ala	

			•
460	465	5	470
tcc cca gga tgg ggc Ser Pro Gly Trp Gly			
475	480	485	GIN NIS Deu
atc gtg tcc ttg ctg Ile Val Ser Leu Leu			
490	495	500	•
acc tct att tagattto	ca ataatcagag a	agtgtaaaa taaaaca	gat tatattgtaa 1600
505			
aaaaaaaaaa aaaaaa			1616
<210> 10			
<211> 507			
<212> PRT			
<213> Homo sapiens			
<400> 10			
Met Asp Gly Thr Glu	Thr Arg Gln Ar	g Arg Leu Asp Ser	Cys Gly Lys
1 5		10	15
Pro Gly Glu Leu Gly 20		o Leu Ser Thr Gly	Gly Leu Pro
Val Ala Ser Glu Asp	_	-	•
. 35	40	45	
Thr Pro Lys Pro Leu	Glu Thr Glu Pr	o Ser Arg Glu Th	Ala Trp Ser
50	55	60	
Ile Gly Leu Gln Val	Thr Val Pro Ph	ne Met Phe Ala Gly	y Leu Gly Leu
65	70	75	. <u>-</u> 80
	. •		
Ser Trp Ala Gly Met			
85		90	95



		•	100		٠.			105					110		
Lys	Gly	Asn 115	Leu	Glu	Met	Thr	Leu 120	Ala	Ser	Arg	Leu	Ser 125	Thr	Ala	Ala
Asn	Thr 130	Gly	Gln	Ile	Asp	Asp 135	Pro	Gln	Glu	Gln	His 140	Arg	Val	Ile	Ser
Ser 145	Asn	Leu	Ala	Leu	11e 150	Gln	Val	Gln	Ala	Thr 155	Val	Val	Gly	Leu	Leu 160
Ala	Ala	Val	Ala	Ala 165	Leu	Leu	Leu	Gly	Val 170	Val	Ser	Arg	Glu	Glu 175	Val
Asp	Val	Ala	Lys 180	Val	Glu	Leu	Leu	Cys 185	Ala	Ser	Ser	Val	Leu 190	Thr	Ala
Phe	Leu	Ala 195	Ala	Phe	Ala	Leu	Gly 200	Val	Leu	Met	Val	Cys 205	Ile	Val	Ile
Gly	Ala 210	Arg	Lys	Leu	Gly	Val 215	Asn	Pro	Asp	Asn	11e 220	Ala	Thr	Pro	Ile
Ala .225	Ala	Ser	Leu	Gly	Asp 230	Leu	Ile	Thr	Leu	Ser 235	Ile	Leu	Ala	Leu	Val 240
Ser	Ser	Phe	Phe	Tyr 245	Arg	His	Lys	Asp	Ser 250	Arg	Tyr	Leu	Thr	Pro 255	Leu
Val	Cys	Leu	Ser 260	Phe	Ala	Ala	Leu	Thr 265	Pro	Vạl	Trp	Val	Leu 270	Ile	Ala
Lys	Gln	Ser 275	Pro	Pro	Ile	Val	Lys 280		Leu	Lys	Phe	Gly 285		Phe	Pro
Ile	Ile 290		Ala	Met	Val	Ile 295		Ser	Phe	Gly	Gly 300		Ile	Leu	Ser

Val Glu Val Lys Asp Leu Leu Thr Leu Val Pro Pro Leu Val Gly Leu

WO 2004/016637



ъуs 305	INT	val	ser	гÀз	310	GIN	туг	тÀз	СТĀ	315	AIA	ıre	ьùе	TNT	320
Val	Ile	Cys	Gly	Val 325	Gly	Gly	Asn	Leu	Val 330	Ala	Ile	Gln	Thr	Ser 335	Arg
Ile	Ser	Thr	Tyr 340	Leu	His	Met	Trp	Ser 345	Ala	Pro	Gly	Val	Ьец 350	Pro	Leu
Gln	Met	Lys 355	Lys	Phe	Trp	Pro	Asn 360	Pro	Cys	Ser	Thr	Phe 365	Cys	Thr	Ser
Glu	Ile 370	Asn	Ser	Met	Ser	Ala 375	Arg	Val	Leu	Leu	Leu 380	Leu	Val	Val	Pro
Gly 385	His	Leu	Ile	Phe	Phe 390	Tyr	Ile	Ile	Tyr	Leu 395	Val	Glu	Gly	Gln	Ser 400
Val	Ile	Asn	Ser	Gln 405	Thr	Phe	Val	Val	Leu 410	Tyr	Leu	Leu		Gly 415	Leu
Ile	Gln	Val	Thr 420	Ile	Leu	Leu	Tyr	Leu 425	Ala	Glu	Val	Met	Val 430	Arg	Leu
Thr	Trp	His 435	Gln	Ala	Leu	Asp	Pro 440	Asp	Asn	His	Cys	Ile 445	Pro	Tyr	Leu
Thr	Gly 450	Leu	Gly	Asp	Leu	Leu 455	Gly	Ser	Ser	Ser	Val 460		His	Thr	Ala
Ala 465	Val	Pro	Arg	Arg <sub>.</sub>	Cys 470	Thr	Ala	Ser	Pro	Gly 475	Trp	Gly	Leu		Gln 480
Pro	Phe	Ile	Суз	Thr 485		His	Leu	Ile	Val 490		Leu	Leu ·	Ser	Phe 495	Tyr
Phe	Pro	Phe	Cys		Leu	Ala	Lys	Thr		Ile					

#### INTERNATIONAL SEARCH REPORT

nternational application No. PR03/01655

A. CLAS	A. CLASSIFICATION OF SUBJECT MATTER								
IPC7	IPC7 C07H 21/00								
According to Ir	nternational Patent Classification (IPC) or to both nation	nal classification and IPC							
	OS SEARCHED								
Minimum docu	mentation searched (classification system followed by	ciassification symbols)	· ·						
H C / CU/11 2	<i>1</i> /00								
Documentation	Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched								
Electronic data	base consulted during the intertnational search (name of	of data hace and where practicable search term	as used)						
	ndant GenBank+PDB+SwissProt+PIR+PRF, Korean Pa								
C. DOCUM	IENTS CONSIDERED TO BE RELEVANT								
Category*	Citation of document, with indication, where appr	ropriate, of the relevant passages	Relevant to claim No.						
X, P	Mammalian Gene Collection (MGC) Program Team		1-17, 27-28, 34						
	15,000 full-length human and mouse cDNA sequence of Science, U. S. A. (24. Dec. 2002) Vol. 99(26):1689		•						
X, P	. TROELS WABAKKEN, EDITH RIAN, MARIT KV	EINE, HANS-CHRISTIAN AASHEIM 'The	1, 10-20, 27-28, 34						
	human solute carrier SLC41A1 belongs to a novel cul	caryotic subfamily with homology to	•						
ļ	prokaryotic MgtE Mg2+ transporters' In Biochem. Biophys. Res. Commun. (Jul. 2003) Vol. 306(3):718-724								
x	Accession No. AK093380 (15. Jul. 2002)		1, 4-6, 11-14, 18-20,						
	27-28								
x	Accession No. BAB14241 (1. Aug. 2002)								
	. 27-28, 34								
x	X Accession No. BAC03717 (15. Jul. 2002)								
	27-28, 34								
Further	documents are listed in the continuation of Box C.	See patent family annex.							
	ategories of cited documents:	"T" later document published after the internation							
to be of pa	"A" document defining the general state of the art which is not considered to be of particular relevance date and not in conflict with the application but cited to understand the principle or theory underlying the invention								
1 4									
special re	special reason (as specified) considered to involve an inventive step when the document is								
means	means being obvious to a person skilled in the art								
than the p	"P" document published prior to the international filing date but later "&" document member of the same patent family than the priority date claimed								
Date of the ac	Date of the actual completion of the international search  Date of mailing of the international search report								
0	01 DECEMBER 2003 (01.12.2003) 02 DECEMBER 2003 (02.12.2003)								
	Name and mailing address of the ISA/KR  Authorized officer								
	Korean Intellectual Property Office 920 Dunsan-dong, Seo-gu, Daejeon 302-701, Republic of Korea	LEE, Yoon Won							
A13 A	csimile No. 82-42-472-7140  Telephone No. 82-42-481-5852								

Box I	Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)
This intern	ational search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:
b	claims Nos.: 29-33 ecause they relate to subject matter not required to be searched by this Authority, namely: Claims 29 to 33 pertain to methods for treatment of the human or animal body by therapy, as well as diagnostic methods, and
	relate a subject matter which this International Searching Authority is not required, under Article 17(2)(a)(i) of the PCT and Rule 39.1(iv) of the Regulations under the PCT, to search.
ا لــا ئ	Claims Nos.:  lecause they relate to part of the international application that do not comply with the prescribed requirements to such an eaningful international search can be carried out, specifically:
· · · · · · · · · · · · · · · · · · ·	Claims Nos.:  because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).
Box II	Observations where unity of invention is lacking (Continuation of item 2 of first sheet)
This Inter	national Search Authority found multiple inventions in this international application, as follows:
	·
1 11	As all required addtional search fees were timely paid by the applicant, this international search report covers all searchable claims.
	As all searchable claims could be established without effort justifying an additional fee, this Authority did not invite payment of any addition fee.
	As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
	No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
Remark	on Protest
	No protest accompanied the payment of additional search fees.